

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:31:39 ; Search time 77.7712 Seconds  
(without alignments)  
458.180 Million cell updates/sec

Title: US-10-797-821-29  
Perfect score: 2104  
Sequence: 1 MKKRILSAVLVSGVTLSSAT.....SIGNRGWFNPGSVSYIYPN 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1057.5	50.3	399	2	US-09-107-433-3230
2	1036.5	50.2	392	2	US-09-583-110-4374
3	670	31.8	461	1	US-08-186-222-2
4	587	27.9	525	2	US-09-107-532A-5095
5	562	26.7	210	2	US-09-222-938A-67
6	524.5	24.9	497	2	US-09-134-000C-5990
7	507	24.1	449	2	US-09-071-035-482
8	507	24.1	449	2	US-10-206-576-482
9	507	24.1	450	2	US-09-134-000C-5714
10	458	21.8	422	2	US-09-071-035-484
11	458	21.8	422	2	US-10-206-576-484
12	278	13.2	469	2	US-09-489-039A-13565
13	238.5	11.3	2310	2	US-09-874-923-120
14	233	11.1	257	2	US-09-710-279-3244
15	233	11.1	264	2	US-09-134-001C-5035
16	233	11.1	267	2	US-09-134-001C-4539
17	226.5	10.8	477	2	US-09-902-540-11649
18	220.5	10.5	1236	2	US-09-769-787-109
19	219	10.4	468	2	US-09-328-352-6321
20	217	10.3	270	2	US-09-134-001C-5441
21	215	10.2	149	2	US-09-710-279-1682
22	215	10.2	157	2	US-09-710-279-2870
23	211.5	10.1	610	2	US-09-336-447A-11
24	211.5	10.1	610	2	US-09-952-267B-11
25	210.5	10.0	440	2	US-08-302-756E-35
26	210	10.0	266	2	US-09-134-001C-5453
27	210	10.0	639	2	US-09-902-540-14908

28	207	9.8	679	2	US-09-489-039A-12307	Sequence 12307, A
29	206	9.8	1070	2	US-09-902-540-13861	Sequence 13861, A
30	202.5	9.6	1566	1	US-08-687-956A-23	Sequence 23, Appl
31	202	9.6	1024	2	US-09-270-767-44973	Sequence 44973, A
32	200.5	9.5	655	2	US-09-902-540-10005	Sequence 10005, A
33	200	9.5	955	1	US-08-006-876B-1	Sequence 1, Appli
34	200	9.5	955	1	US-08-282-845-2	Sequence 2, Appli
35	200	9.5	955	1	US-08-428-414A-3	Sequence 3, Appli
36	200	9.5	955	4	PCT-US94-00324-1	Sequence 1, Appli
37	198	9.4	928	2	US-09-134-000C-6590	Sequence 6590, Ap
38	195.5	9.3	2101	1	US-08-466-390-4	Sequence 4, Appli
39	195.5	9.3	2101	1	US-08-470-950-4	Sequence 4, Appli
40	195.5	9.3	2101	1	US-08-467-781-4	Sequence 4, Appli
41	195.5	9.3	2101	1	US-08-195-487-4	Sequence 4, Appli
42	195.5	9.3	2101	1	US-08-483-924-4	Sequence 4, Appli
43	195.5	9.3	2101	2	US-09-452-294-1	Sequence 1, Appli
44	195.5	9.3	2101	2	US-09-296-662-32	Sequence 32, Appl
45	195.5	9.3	2101	4	PCT-US93-06160-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-09-107-433-3230  
; Sequence 3230, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
THERAPEUTICS

NUMBER OF SEQUENCES: 5206  
CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:  
MEDIUM TYPE: CD/ROM ISO9660  
COMPUTER: <Unknown>

OPERATING SYSTEM: <Unknown>

SOFTWARE: <Unknown>

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,433

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/ 085131

FILING DATE: May 12, 1998

APPLICATION NUMBER: 60/051553

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-011

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 3230:

SEQUENCE CHARACTERISTICS:

LENGTH: 399 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Streptococcus pneumoniae

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1...399

SEQUENCE DESCRIPTION: SEQ ID NO: 3230:

US-09-107-433-3230

Query Match 50.3%; Score 1057.5; DB 2; Length 399;  
Best Local Similarity 50.8%; Pred. No. 1.2e-69;  
Matches 221; Conservative 74; Mismatches 89; Indels 51; Gaps 6;

Qy 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQTASQDSKINNLTAAQQAAQAVNTIQG 60  
Db 8 MKKKILASLLSVWVSQAVLTTAHTTDDKIAQDNKISNLTAAQQAAQAVNTIQG 67

Qy 61 QVSALOTQQAELQENORLEAQSATLGGQIQTSSKIVARNESLKKQARSQAQSNATYS 120  
Db 68 QVSAIQAEQSNLOAENDRLQAESEKLEGEITELSKNIVSRNQSLEKQARSQAQNGAVTSY 127

Qy 121 INALINSKVSDAINRVSAREVVSANEKMLQOEODKAAVEKQOQENQAINVAAQOE 180  
Db 128 INTIVNSKSITEAISRVAAEISEVSNKMLQOEODKAAVEKQOQENQAINVAAQOE 187

Qy 181 TIAQNTNALNTQQAELQENORLEAQSATLGGQIQTSSKIVARNESLKKQARSQAQSNATYS 240  
Db 188 KLADDAQALTTKQNELKAAELSLAAEKATAEKASLEKQAAAEAEARAAVAEAAAYKE 247

Qy 241 KAAAEAKALQEQAAQAVANNNTQATDASDQQAADNTQAAQTGDSSTEQSAAQAVNNS 300  
Db 248 KRASQ-----QOSVLASANTNLT-----AQVQAVSESAAA----- 277

Qy 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGCTGKSLAPWVGNVWNGGOWAA 360  
Db 278 -----PVRKVRPTYS-----TNASSYPIGECTGKSLAPWVGNVWNGGOWAA 322

Qy 361 SAAAAGRYVSGTSPSAGAVAVMNDGGYGHVAVYTVGQV-GQIQVQEAANYAGNQSIGNYRGW 419  
Db 323 SAAAAGRTGSTPQVGAICWMDGGYGHVAVVTVAVESTTRIQVSESNYAGNRTIGNHRGW 382

Qy 420 FNP-----GSVSIYI 429  
Db 383 FNPTTSEGFTVIYI 397

RESULT 2

US-09-583-110-4374  
; Sequence 4374, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 4374  
; LENGTH: 392  
; TYPE: PR1  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-4374

Query Match 50.2%; Score 1056.5; DB 2; Length 392;  
Best Local Similarity 50.8%; Pred. No. 1.4e-69;  
Matches 221; Conservative 73; Mismatches 90; Indels 51; Gaps 6;

Qy 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQTASQDSKINNLTAAQQAAQAVNTIQG 60  
Db 1 MKKKILASLLSVWVSQAVLTTAHTTDDKIAQDNKISNLTAAQQAAQAVNTIQG 60

Qy 61 QVSALOTQQAELQENORLEAQSATLGGQIQTSSKIVARNESLKKQARSQAQSNATYS 120  
Db 68 QVSAIQAEQSNLOAENDRLQAESEKLEGEITELSKNIVSRNQSLEKQARSQAQNGAVTSY 127

Db 61 QVSAIQAEQSNLOAENDRLQAESEKLEGEITELSKNIVSRNQSLEKQARSQAQNGAVTSY 120

Qy 121 INALINSKVSDAINRVSAREVVSANEKMLQOEODKAAVEKQOQENQAINVAAQOE 180  
Db 128 INTIVNSKSITEAISRVAAEISEVSNKMLQOEODKAAVEKQOQENQAINVAAQOE 180

Qy 181 TIAQNTNALNTQQAELQENORLEAQSATLGGQIQTSSKIVARNESLKKQARSQAQSNATYS 240  
Db 188 KLADDAQALTTKQNELKAAELSLAAEKATAEKASLEKQAAAEAEARAAVAEAAAYKE 240

Qy 241 KAAAEAKALQEQAAQAVANNNTQATDASDQQAADNTQAAQTGDSSTEQSAAQAVNNS 300  
Db 248 KRASQ-----QOSVLASANTNLT-----AQVQAVSESAAA----- 270

Qy 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGCTGKSLAPWVGNVWNGGOWAA 360  
Db 271 -----PVRKVRPTYS-----TNASSYPIGECTGKSLAPWVGNVWNGGOWAA 315

Qy 361 SAAAAGRYVSGTSPSAGAVAVMNDGGYGHVAVYTVGQV-GQIQVQEAANYAGNQSIGNYRGW 419  
Db 316 SAAAAGRTGSTPQVGAICWMDGGYGHVAVVTVAVESTTRIQVSESNYAGNRTIGNHRGW 375

Qy 420 FNP-----GSVSIYI 429  
Db 376 FNPTTSEGFTVIYI 390

RESULT 3

US-08-186-222-2  
; Sequence 2, Application US/08186222  
; Patent No. 5559007  
; GENERAL INFORMATION:  
; APPLICANT: Suri, Bruno  
; APPLICANT: Schmitz, Albert  
; TITLE OF INVENTION: Bacterial Vectors  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/186,222  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/672,205  
; FILING DATE: 19-MAR-1991  
; APPLICATION NUMBER: GB 9006400.7  
; FILING DATE: 22-MAR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Villamizar, JoAnn  
; REGISTRATION NUMBER: 30,598  
; REFERENCE/DOCKET NUMBER: 4-17994/A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (914)785-7121  
; TELEFAX: (914)347-5769  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 461 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-186-222-2

Query Match 31.8%; Score 670; DB 1; Length 461;



[illegible]

```
; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-67

Query Match          26.7%; Score 562; DB 2; Length 210;
Best Local Similarity 55.2%; Pred. No. 1e-33;
Matches 116; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASODSKINNLTAAQQAAQAQVNTIQ 60
Db 1 MKKKILASLLSTVWVSQVAVLTTHAETTDKIAADNKISLNTAAQQAAQAQVNTIQ 60

Qy 61 QVSALOTQQAELQAEORLEAQSATLQQOIQTLSSKIVARNESLKKQARSQAQSNAAATSY 120
Db 61 QVSAIQAEQSNLQAEORLEAQSATLQQOIQTLSSKIVARNESLKKQARSQAQSNAAATSY 120

Qy 121 INAIINSKVSDAINRVSAITREVVSAANEKMLQOEQDQAAVEQKQAEQQAINTVAANO 180
Db 121 INTIVNSKISITAEISRVAAAMEIVSANNKMLEQKADKKAISEKQVANNDAINTVIANQ 180

Qy 181 TIAQNTNALNTQQAQLEAAQLNLQAEULTTA 210
Db 181 KLADDAQALITTKQELKAAELSLAAEKATS 210

RESULT 6
US-09-134-000C-5990
; Sequence 5990, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5990
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5990

Query Match          24.9%; Score 524.5; DB 2; Length 497;
Best Local Similarity 31.2%; Pred. No. 1.7e-30;
Matches 151; Conservative 82; Mismatches 176; Indels 75; Gaps 12;

Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASODSKINNLTAAQQAAQAQVNTIQ 60
Db 16 LKKSIVLSALMVCSTILTSVALPSAFADEYDTKIQODQKINALTSQMSDAEAKVAIEN 75

Qy 61 QVSALOTQQAELQAEORLEAQSATLQQOIQTLSSKIVARNESLKKQARSQAQSNAAATSY 120
Db 76 DMVETAKQIDTLTAKKNKLSSEVSKLYSEISDLNVRIQREVQMTKQARDVQVNGQSDSI 135

Qy 121 INAIINSKVSDAINRVSAITREVVSAANEKMLQOEQDQAAVEQKQAEQQAINTVAANO 180
Db 136 IDAVLDADSADAIQVAVTQVAVTQVAVTQVAVTQVAVTQVAVTQVAVTQVAVTQVAV 195

Qy 181 TIAQNTNALNTQQAQLEAAQLNLQAEULTTAQDQKATLV-----AOKAAAEAAARQAAAA- 234
Db 196 ELMDKTESLTKLIQAEVAKNDLQAEQSEQKQDGFQKQKAEKRLAEQARQAAK 255

Qy 235 -----QAAAEKAAAEKALQEAQAAQVAAANNNTQATDSDQQAADNTQAAQTGDSTE 290
Db 256 KAEQEAQAAQAAQAAQ-KAAAEQAKATKAA---NEAAASAAAEKAA---TPVVESSTTTE 307

Qy 291 QSAQAQVNNSDQES-----TTATAEQP-----SASSASTAA 321
Db 308 STTTQETTTSTETESVVTTPVAAAPAEKEKEVPVTPNTPPTPEKGNKAPGNGVTSQKAA 367
```

```
; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-67

Query Match          26.7%; Score 562; DB 2; Length 210;
Best Local Similarity 55.2%; Pred. No. 1e-33;
Matches 116; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASODSKINNLTAAQQAAQAQVNTIQ 60
Db 1 MKKKILASLLSTVWVSQVAVLTTHAETTDKIAADNKISLNTAAQQAAQAQVNTIQ 60

Qy 61 QVSALOTQQAELQAEORLEAQSATLQQOIQTLSSKIVARNESLKKQARSQAQSNAAATSY 120
Db 61 QVSAIQAEQSNLQAEORLEAQSATLQQOIQTLSSKIVARNESLKKQARSQAQSNAAATSY 120

Qy 121 INAIINSKVSDAINRVSAITREVVSAANEKMLQOEQDQAAVEQKQAEQQAINTVAANO 180
Db 121 INTIVNSKISITAEISRVAAAMEIVSANNKMLEQKADKKAISEKQVANNDAINTVIANQ 180

Qy 181 TIAQNTNALNTQQAQLEAAQLNLQAEULTTA 210
Db 181 KLADDAQALITTKQELKAAELSLAAEKATS 210

RESULT 7
US-09-071-035-482
; Sequence 482, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 482:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-071-035-482

Query Match          24.1%; Score 507; DB 2; Length 449;
Best Local Similarity 31.6%; Pred. No. 2.9e-29;
Matches 142; Conservative 88; Mismatches 175; Indels 44; Gaps 12;

Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASODSKINNLTAAQQAAQAQVNTIQ 60
Db 1 VKKELFASVLLCSLTLSAITSIALADNVDKKIBEKQEISSLKAKQGLASQVSSLEA 60

Qy 61 QVSALOTQQAELQAEORLEAQSATLQQOIQTLSSKIVARNESLKKQARSQAQSNAAATSY 120
Db 61 EVSSVFDESMALREQKTLKAKSEQLQOEITNLNQRNEIKRNEAIKQARDVQVNGOSTTM 120

Qy 121 INAIINSKVSDAINRVSAITREVVSAANEKMLQOEQDQAAVEQKQAEQQAINTVAANO 180
Db 121 LDVLDADSADAIQVAVTQVAVTQVAVTQVAVTQVAVTQVAVTQVAVTQVAVTQVAV 180

Qy 181 TIAQNTNALNTQQAQLEAAQLNLQAEULTTAQDQKATLVAAQKAAAE-EAARQAAAAQAAAE 239
```

181	Db	:   : : : : :   : : : :	ELETKRQDLISKQSELNWKSLAQSSAESSKAGLEKQKAABEQRLAAEQKAAAE	240
240	QY	-----AKAAAEAKALOEQAQAQVAANNNTQ-----ATDASDQAAA-----AADNTQAAQ	284	
241	Db	:   : : : : :   : : : :	KAKQAAAKPAKAEVKAEAPVASSSTTEAQAPASSSSATESSTQQTTETTPSTDNSATEN	300
285	QY	-----TGDSTEQSAQAQV-----NNSDQESTTATEAQPSASSASTAAVAANTSSANTYPAGQCT	338	
301	Db	:   : : : : :   : : : :	TGSSSSSEQPVQPTTPSDNGNNGGQTGGCTVTPTEPTPAPSDPTINALNLVLRSLG---	357
339	QY	GWVKSLAPWNYNGNGGQWAAASAAAAGYRVGSTPSAGAVAWNDGGYGHVAYVTGVGGG	398	
358	Db	-----LRPV-----WDAGLAASATARAQAOVERAGGIPINDH-----WSRG-----DEVIAIMWAPGN	403	
399	QY	QIQV-----QEANYAGNQSIGNYRGW-FNPG	423	
404	Db	:   : : : : :   : : : :	SVIMAWNYETNMVTSAGSG-HRDWEINPG	431

```

RESULT 8
US-10-206-576-482
; Sequence 482, Application US/10206576
; Patent No. 6913907
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 497
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
;

```

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-R  
COMPUTER: Dell Latitude  
OPERATING SYSTEM: Windows 98  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/206,576  
FILING DATE: 29-Jul-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/071,035  
FILING DATE: 1998-05-04  
APPLICATION NUMBER: US 60/046,655  
FILING DATE: 1997-05-16  
APPLICATION NUMBER: US 60/044,031  
FILING DATE: 1997-05-06  
APPLICATION NUMBER: US 60/066,009  
FILING DATE: 1997-11-14  
ATTORNEY/AGENT INFORMATION:  
NAME: Hyman, Mark J.  
REGISTRATION NUMBER: 46,789  
REFERENCE/DOCKET NUMBER: PB369P1D1  
INFORMATION FOR SEQ ID NO: 482:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 449 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 482:  
US-10-206-576-482

[illegible]

Db	1	VKKRLFASVLLCSUTLSA	ITPSTALADNVDDK	KTIEKNQEISSLKAKOGD	LASQVSSLEA	60								
Qy	61	QVSALOTQQAELQAE	NRLEAQA	SATLGQOI	QITLSSKIVARNESL	KQQAARSQA	KNSAATSY	120						
Db	61	EVSSVPDES	WALREKOTL	KAKSELOQE	ITNLNQR	IEKRN	EALKNQARDV	VQVNGOSTTM	120					
Qy	121	INAIINSKYS	DAINRYS	AIREVVSANE	KMLQQOQE	QDKAAVEOR	KQOENQAA	INTVAANQE	180					
Db	121	LDAVLDADS	VADALS	RVQAVSTI	VSANNDLW	QQQKEDQ	AAVVDKKA	ENKEKKVKOLEATEA	180					
Qy	181	TIATQNTAL	NTQQAQL	EAAQNL	QLAEIT	TADQKATL	VAAKAAE	-EAA	QAAAAQAAAAE	239				
Db	181	ELETKRQD	LLSKQSEL	NVMKAS	LAL	EQSSAESS	KAGLEKQ	KAAAEAEQ	ARLAAEQKAAAE	240				
Qy	240	-AKAAAE	AKALQEQAAQ	AAVANN	NTQ-----	ATDASDQAA	-----	AA	NTQAAQ	284				
Db	241	KAKQAAAK	PAKAEV	KAEAPV	ASSSTTEA	QAPASS	SSSATES	STQTQT	TETTPSTD	NSATEN	300			
Qy	285	TGDS	TGESA	AAQV-----	NNSQD	ESTTATE	QAQSPASS	ASASTAA	VAANTSS	ANTYPAG	OCT	338		
Db	301	TGSSSS	EQPQPT	TPSD	NGNNGQ	TGGT	VTPT	PEPT	PAPS	ADPT	INALN	LRQSLG	---	357
Qy	339	WGVKSL	APWTGV	WNGCGQ	WAASAAA	AGYRV	GSTPS	SAGAV	VWNDGG	YGHV	YVYTG	VOGG	398	
Db	358	-----	LRPVV	---WD	GLAAS	ATARA	QVEAGG	IPNDH	---WS	RG-	-DEVIA	IMWAPGN	403	
Qy	399	QIQV	---QE	ANYAG	NQSIG	NYRGW	-FNPG	423						
Db	404	SVTMAW	NETNM	TYTAS	GS	-HRD	WEINPG	431						

```

RESULT 9
US-09-134-000C-5714
; Sequence 5714, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 5714
; LENGTH: 450
; TYPE: prt
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5714

```

Query Match	24.1%	Score 507;	DB 2;	Length 450;
Best Local Similarity	31.6%;	Pred. No. 2.9e-29;		
Matches 142;	Conservative 88;	Mismatches 175;	Indels 44;	Gaps 12;
Qy	1	MKKRIISAVLVSGVTISSLATTL	SAVKADDDAQI	IASODSKINNLTAQOQAAQAQVNTIQG 60
Db	2	VKKRLPASVLLCSLTLSA	ITPSIALADNVDKLIEEKQ	QEISSLKAKQGDLSQVSSLEA 61
Qy	61	QVSGALQTQQAELQAEQNR	LEAQSAATLGGQITQLTSSK	IVARNESLKQARSQAQKSNATSY 120
Db	62	EVSSVFDSESMALREQTK	LKAKSEQLOEITNLNQR	LIEKNEAIAKQARDVQVNGQSTTM 121
Qy	121	INAIINSKSVSDAINRV	SATREVVSNEMKMLQOQ	QDQKAAVEQKQSNQAAINTVAANQE 180
Db	122	LDAVLDA	DSVADAI	SRVQAVSTIVSANNDLMQOQKEDQAVVDKKAENEKVKQLEATEA 181
Qy	181	TIQAQNTVALNTQQAQ	LEAAQLNQLAELTTHAQDQ	KATLVAQKAAAE-EAARQAAAAQAQAAE 239
Db	182	ELETKRQDLLSKSGEL	NVMPKASLALBOSSAES	KAGLEKQKAAAEABQARLLAAEQKAAAE 241
Qy	240	-AKAAAREAKALQ	BQAAQVAAANNNTQ-	-----ATDASDQAAA-----AADNTQAAQ 284





[illegible]

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:14:04 ; Search time 252.865 Seconds  
(without alignments)  
712.176 Million cell updates/sec

Title: US-10-797-821-29  
Perfect score: 2104  
Sequence: 1 MKKRILSAVLVSGVTLLSAT.....SIGNRGWFGNSVSIYPN 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA\_Main.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2104	100.0	431	4	US-10-383-930-29
2	2104	100.0	431	5	US-10-797-821-29
3	2100	99.8	431	4	US-10-383-930-30
4	2100	99.8	431	5	US-10-797-821-30
5	2092	99.4	431	4	US-10-383-930-33
6	2092	99.4	431	5	US-10-797-821-33
7	2079.5	98.8	432	4	US-10-383-930-31
8	2079.5	98.8	432	5	US-10-797-821-31
9	2070.5	98.4	432	4	US-10-383-930-32
10	2070.5	98.4	432	5	US-10-797-821-32
11	1085.5	51.6	398	5	US-10-474-792-600
12	1057.5	50.3	392	5	US-10-472-928-4652
13	1057.5	50.3	399	5	US-10-617-320-3230
14	591	28.1	524	4	US-10-282-122A-57658
15	562	26.7	210	4	US-10-154-251-67
16	507	24.1	449	3	US-09-071-035-482
17	507	24.1	449	4	US-10-206-576-482
18	507	24.1	449	5	US-10-912-362-482
19	458	21.8	422	3	US-09-071-035-484
20	458	21.8	422	4	US-10-206-576-484
21	458	21.8	422	5	US-10-912-362-484
22	333.5	15.9	630	5	US-10-494-674-6
23	327.5	15.6	600	3	US-09-738-626-5197
24	298	14.2	609	5	US-10-501-282-3184
25	295	14.0	440	5	US-10-501-282-3056
26	292.5	13.9	422	5	US-10-501-282-3054
27	283.5	13.5	377	5	US-10-501-282-3052

28	263	12.5	428	4	US-10-282-122A-55748	Sequence 55748, A
29	245.5	11.7	484	5	US-10-282-122A-50526	Sequence 50526, A
30	245.5	11.7	558	4	US-10-732-923-3295	Sequence 3295, Ap
31	238.5	11.3	2310	3	US-09-874-923-120	Sequence 120, App
32	238.5	11.3	2310	3	US-09-991-496-120	Sequence 120, App
33	238.5	11.3	2310	3	US-09-820-843A-114	Sequence 114, App
34	237.5	11.3	359	5	US-10-850-274-86	Sequence 86, Appl
35	233	11.1	257	5	US-10-470-048B-362	Sequence 362, App
36	233	11.1	264	4	US-10-724-972A-6539	Sequence 6539, Ap
37	233	11.1	267	4	US-10-724-972A-5110	Sequence 5110, Ap
38	233	11.1	2354	3	US-09-820-843A-113	Sequence 113, App
39	231	11.0	1545	5	US-10-732-923-2248	Sequence 2248, Ap
40	230.5	11.0	421	4	US-10-282-122A-56483	Sequence 56483, A
41	229.5	10.9	1463	3	US-09-971-536-69	Sequence 69, Appl
42	225.5	10.7	166	5	US-10-470-048B-73	Sequence 73, Appl
43	222.5	10.6	267	5	US-10-470-048B-74	Sequence 74, Appl
44	222	10.6	323	4	US-10-282-122A-59321	Sequence 59321, A
45	221.5	10.5	265	5	US-10-470-048B-89	Sequence 89, Appl

ALIGNMENTS

RESULT 1  
US-10-383-930-29  
; Sequence 29, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 29  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-29

Query Match	100.0%;	Score 2104;	DB 4;	Length 431;
Best Local Similarity	100.0%;	Pred. No. 1e-119;		
Matches 431;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MKKRILSAVLVSGVTLLSSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG	60	
Db	1	MKKRILSAVLVSGVTLLSSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG	60	
Qy	61	QVSALOTQQAELQAEQNRLEAQAQATLGGQIQTQLSSKIVARNESLKQQAARSQAQKSNAATSY	120	
Db	61	QVSALOTQQAELQAEQNRLEAQAQATLGGQIQTQLSSKIVARNESLKQQAARSQAQKSNAATSY	120	
Qy	121	INAIINSKVSVDAINRVSAIREVVSANERKMLQOEODKAAVEQKQEQENQAIAINTVAANQE	180	
Db	121	INAIINSKVSVDAINRVSAIREVVSANERKMLQOEODKAAVEQKQEQENQAIAINTVAANQE	180	
Qy	181	TTAQTNTNALNTQQAQLEAAQLNLQAEELTTAODQKATLVAQKAAAEAAARQAAAAQAAAA	240	
Db	181	TTAQTNTNALNTQQAQLEAAQLNLQAEELTTAODQKATLVAQKAAAEAAARQAAAAQAAAA	240	
Qy	241	KAAAEAKALQEQAAQAQVAANNNTQTATDASDQAAAAADNTQAAQTGDTGSTGSAQAQVYNS	300	
Db	241	KAAAEAKALQEQAAQAQVAANNNTQTATDASDQAAAAADNTQAAQTGDTGSTGSAQAQVYNS	300	
Qy	301	QDSESTTATEAQPSASSASTAAVAANTSSANTYPAGOCCTGCVKSLAPWVGNWNGGOWAA	360	
Db	301	QDSESTTATEAQPSASSASTAAVAANTSSANTYPAGOCCTGCVKSLAPWVGNWNGGOWAA	360	

QY 361 SAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGGGQIQVQEQVAGNOSIGNYRGWF 420  
Db 361 SAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGGGQIQVQEQVAGNOSIGNYRGWF 420

QY 421 NPGSVSYIYPN 431  
Db 421 NPGSVSYIYPN 431

RESULT 2  
US-10-797-821-29  
; Sequence 29, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 29  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-29

Query Match 100.0%; Score 2104; DB 5; Length 431;  
Best Local Similarity 100.0%; Pred. No. 1e-119;  
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60  
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

QY 61 QVSALOTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNATSY 120  
Db 61 QVSALOTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAIAINTVAANQE 180  
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAIAINTVAANQE 180

QY 181 TIAQNTNALNTQQAQLAEALQNLQAEITTAQDQKATLVAQKAAAEAEARQAAAAQAAEA 240  
Db 181 TIAQNTNALNTQQAQLAEALQNLQAEITTAQDQKATLVAQKAAAEAEARQAAAAQAAEA 240

QY 241 KAAAEAKALQEQAAQVAANNNTQATDASDQQAADNTQAAQTGDSSTQSAQAQVNN 300  
Db 241 KAAAEAKALQEQAAQVAANNNTQATDASDQQAADNTQAAQTGDSSTQSAQAQVNN 300

QY 301 DQESTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360  
Db 301 DQESTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360

QY 361 SAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGGGQIQVQEQVAGNOSIGNYRGWF 420  
Db 361 SAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGGGQIQVQEQVAGNOSIGNYRGWF 420

QY 421 NPGSVSYIYPN 431  
Db 421 NPGSVSYIYPN 431

RESULT 3  
US-10-383-930-30  
; Sequence 30, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 30  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-30

Query Match 99.8%; Score 2100; DB 4; Length 431;  
Best Local Similarity 99.8%; Pred. No. 1.8e-119;  
Matches 430; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60  
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

QY 61 QVSALOTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNATSY 120  
Db 61 QVSALOTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAIAINTVAANQE 180  
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAIAINTVAANQE 180

QY 181 TIAQNTNALNTQQAQLAEALQNLQAEITTAQDQKATLVAQKAAAEAEARQAAAAQAAEA 240  
Db 181 TIAQNTNALNTQQAQLAEALQNLQAEITTAQDQKATLVAQKAAAEAEARQAAAAQAAEA 240

QY 241 KAAAEAKALQEQAAQVAANNNTQATDASDQQAADNTQAAQTGDSSTQSAQAQVNN 300  
Db 241 KAAAEAKALQEQAAQVAANNNTQATDASDQQAADNTQAAQTGDSSTQSAQAQVNN 300

QY 301 DQESTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360  
Db 301 DQESTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360

QY 361 SAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGGGQIQVQEQVAGNOSIGNYRGWF 420  
Db 361 SAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGGGQIQVQEQVAGNOSIGNYRGWF 420

QY 421 NPGSVSYIYPN 431  
Db 421 NPGSVSYIYPN 431

RESULT 4  
US-10-797-821-30  
; Sequence 30, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens



```
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; PRIOR FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; ORGANISM: Streptococcus mutans
US-10-797-821-30

Query Match          99.8%; Score 2100; DB 5; Length 431;
Best Local Similarity 99.8%; Pred. No. 1.8e-119;
Matches 430; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSSTLTSVAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSSTLTSVAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
Db 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120

Qy 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQEDKAAVEQKQENQAAINTVAANO 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQEDKAAVEQKQENQAAINTVAANO 180

Qy 181 TIAQNTNALNTQQAELQAEALNLQAEALTTAQDQKATLVAQKAAAEAEARQAAAAA 240
Db 181 TIAQNTNALNTQQAELQAEALNLQAEALTTAQDQKATLVAQKAAAEAEARQAAAAA 240

Qy 241 KAAAEAKALQEAQAAQVAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNS 300
Db 241 KAAAEAKALQEAQAAQVAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNS 300

Qy 301 DOESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGOWAA 360
Db 301 DOESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGOWAA 360

Qy 361 SAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYVTGVQGGQIQVQEAANYAGNQSIGNYRGWF 420
Db 361 SAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYVTGVQGGQIQVQEAANYAGNQSIGNYRGWF 420

Qy 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431
```

RESULT 5

```
US-10-383-930-33
; Sequence 33, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
```

```
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; ORGANISM: Streptococcus mutans
US-10-383-930-33

Query Match          99.4%; Score 2092; DB 4; Length 431;
Best Local Similarity 99.3%; Pred. No. 5.6e-119;
Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSSTLTSVAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSSTLTSVAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
Db 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120

Qy 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQEDKAAVEQKQENQAAINTVAANO 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQEDKAAVEQKQENQAAINTVAANO 180

Qy 181 TIAQNTNALNTQQAELQAEALNLQAEALTTAQDQKATLVAQKAAAEAEARQAAAAA 240
Db 181 TIAQNTNALNTQQAELQAEALNLQAEALTTAQDQKATLVAQKAAAEAEARQAAAAA 240

Qy 241 KAAAEAKALQEAQAAQVAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNS 300
Db 241 KAAAEAKALQEAQAAQVAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNS 300

Qy 301 DOESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGOWAA 360
Db 301 DOESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGOWAA 360

Qy 361 SAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYVTGVQGGQIQVQEAANYAGNQSIGNYRGWF 420
Db 361 SAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYVTGVQGGQIQVQEAANYAGNQSIGNYRGWF 420

Qy 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431
```

RESULT 6

```
US-10-797-821-33
; Sequence 33, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
```

```
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-33

Query Match      99.4%; Score 2092; DB 5; Length 431;
Best Local Similarity 99.3%; Pred. No. 5.6e-119;
Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

Qy 61 QVSALOTQQAELQAEQENRLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNATSY 120
Db 61 QVSALOTQQAELQAEQENRLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNATSY 120

Qy 121 INAIINSKVSDAINRVSAIREVVSANRKLQOQEQDKAAVEQKQENQAAINTVAAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANRKLQOQEQDKAAVEQKQENQAAINTVAAANQE 180

Qy 181 TIAONTNLTQQAELQAEQENRLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNATSY 240
Db 181 TIAONTNLTQQAELQAEQENRLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNATSY 240

Qy 241 KAAAEAKALQEAQAQVAA--NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299
Db 241 KAAAEAKALQEAQAQVAA--NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299

Qy 300 SDQESTTATEAQAQVAA--NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 359
Db 300 SDQESTTATEAQAQVAA--NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 359

Qy 360 ASAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVYVTVGGGQIQVQENYAGNOSIGNYRGW 419
Db 360 ASAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVYVTVGGGQIQVQENYAGNOSIGNYRGW 419

Qy 420 FNPGSVSIYYPN 431
Db 420 FNPGSVSIYYPN 431

RESULT 8
US-10-797-821-31
; Sequence 31, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-31

Query Match      98.8%; Score 2079.5; DB 5; Length 432;
Best Local Similarity 98.8%; Pred. No. 3.2e-118;
Matches 427; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

Qy 61 QVSALOTQQAELQAEQENRLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNATSY 120
Db 61 QVSALOTQQAELQAEQENRLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNATSY 120

Qy 121 INAIINSKVSDAINRVSAIREVVSANRKLQOQEQDKAAVEQKQENQAAINTVAAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANRKLQOQEQDKAAVEQKQENQAAINTVAAANQE 180
```

Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANQE 180  
Qy 181 TIAQNTNALNTQQQAQLAAQLNLQAEELTTAQDQKATLVAKAAAEAEARQAAAAQAAAAEA 240  
Db 181 TIAQNTNALNTQQQAQLAAQLNLQAEELTTAQDQKATLVAKAAAEAEARQAAAAQAAAAEA 240  
Qy 241 KAAAEAKALQEQAAQAAQVAA-NNNTQATDASDQQAADNTQAAQTGDSTEQSAQAQAVNN 299  
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSTEQSAQAQAVNN 300  
Qy 300 SDQESTTATEAQPSSASASTAAVAANTSSANTYPAGCTGWGKSLAPWVGNWNGGQWA 359  
Db 301 SDQESTTATEAQPSSASASTAAVAANTSSANTYPAGCTGWGKSLAPWVGNWNGGQWA 360  
Qy 360 ASAAAAGYRGVSTPSAGAVAVWNDGGYGHVAYVTGVOGGQIQVQEAANYAGNQSIGNYRGW 419  
Db 361 ASAAAAGYRGVSTPSAGAVAVWNDGGYGHVAYVTGVOGGQIQVQEAANYAGNQSIGNYRGW 420  
Qy 420 FNPGSVSYIYPN 431  
Db 421 FNPGSVSYIYPN 432

RESULT 9  
US-10-383-930-32  
; Sequence 32, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 32  
; LENGTH: 432  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-32

Query Match 98.4%; Score 2070.5; DB 4; Length 432;  
Best Local Similarity 98.6%; Pred. No. 1.le-117;  
Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Qy 1 MKKRILSAVLVSGVTLSATTL SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60  
Db 1 MKKRILSAVLVSGVTLSATTL SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60  
Qy 61 QVSALQTQQAELQAEQNRLEAQSATLGQQIQTLLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
Db 61 QVSALQTQQAELQAEQNRLEAQSATLGQQIQTLLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
Qy 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANQE 180  
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDKAAVEQKHQENQAAINTVAANQE 180  
Qy 181 TIAQNTNALNTQQQAQLAAQLNLQAEELTTAQDQKATLVAKAAAEAEARQAAAAQAAAAEA 240  
Db 181 TIAQNTNALNTQQQAQLAAQLNLQAEELTTAQDQKATLVAKAAAEAEARQAAAAQAAAAEA 240  
Qy 241 KAAAEAKALQEQAAQAAQVAA-NNNTQATDASDQQAADNTQAAQTGDSTEQSAQAQAVNN 299  
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSTEQSAQAQAVNN 300  
Qy 300 SDQESTTATEAQPSSASASTAAVAANTSSANTYPAGCTGWGKSLAPWVGNWNGGQWA 359  
Db 301 SDQESTTATEAQPSSASASTAAVAANTSSANTYPAGCTGWGKSLAPWVGNWNGGQWA 360  
Qy 360 ASAAAAGYRGVSTPSAGAVAVWNDGGYGHVAYVTGVOGGQIQVQEAANYAGNQSIGNYRGW 419  
Db 361 ASAAAAGYRGVSTPSAGAVAVWNDGGYGHVAYVTGVOGGQIQVQEAANYAGNQSIGNYRGW 420

Db 301 SDQESTTATEAQPSSASASTAAVAANTSSANTYPAGCTGWGKSLAPWVGNWNGGQWA 360  
Qy 360 ASAAAAGYRGVSTPSAGAVAVWNDGGYGHVAYVTGVOGGQIQVQEAANYAGNQSIGNYRGW 419  
Db 361 ASAAAAGYRGVSTPSAGAVAVWNDGGYGHVAYVTGVOGGQIQVQEAANYAGNQSIGNYRGW 420  
Qy 420 FNPGSVSYIYPN 431  
Db 421 FNPGSVSYIYPN 432

RESULT 10  
US-10-797-821-32  
; Sequence 32, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; PRIOR FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 32  
; LENGTH: 432  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-32

Query Match 98.4%; Score 2070.5; DB 5; Length 432;  
Best Local Similarity 98.6%; Pred. No. 1.le-117;  
Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

Qy 1 MKKRILSAVLVSGVTLSATTL SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60  
Db 1 MKKRILSAVLVSGVTLSATTL SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60  
Qy 61 QVSALQTQQAELQAEQNRLEAQSATLGQQIQTLLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
Db 61 QVSALQTQQAELQAEQNRLEAQSATLGQQIQTLLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
Qy 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANQE 180  
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDKAAVEQKHQENQAAINTVAANQE 180  
Qy 181 TIAQNTNALNTQQQAQLAAQLNLQAEELTTAQDQKATLVAKAAAEAEARQAAAAQAAAAEA 240  
Db 181 TIAQNTNALNTQQQAQLAAQLNLQAEELTTAQDQKATLVAKAAAEAEARQAAAAQAAAAEA 240  
Qy 241 KAAAEAKALQEQAAQAAQVAA-NNNTQATDASDQQAADNTQAAQTGDSTEQSAQAQAVNN 299  
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSTEQSAQAQAVNN 300  
Qy 300 SDQESTTATEAQPSSASASTAAVAANTSSANTYPAGCTGWGKSLAPWVGNWNGGQWA 359  
Db 301 SDQESTTATEAQPSSASASTAAVAANTSSANTYPAGCTGWGKSLAPWVGNWNGGQWA 360  
Qy 360 ASAAAAGYRGVSTPSAGAVAVWNDGGYGHVAYVTGVOGGQIQVQEAANYAGNQSIGNYRGW 419  
Db 361 ASAAAAGYRGVSTPSAGAVAVWNDGGYGHVAYVTGVOGGQIQVQEAANYAGNQSIGNYRGW 420

```
Qy 420 FNPQSVSYIPN 431
Db 421 FNPQSVSYIPN 432

RESULT 11
US-10-474-792-600
; Sequence 600, Application US/10474792
; Publication No. US20040236072A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Stephen
; APPLICANT: Zagursky, Robert
; APPLICANT: Nickbarg, Elliot
; APPLICANT: Winter, Lourie
; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES
; FILE REFERENCE: AM 100399
; CURRENT APPLICATION NUMBER: US/10/474, 792
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 674
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 600
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-474-792-600

Query Match 51.6%; Score 1085.5; DB 5; Length 398;
Best Local Similarity 53.7%; Pred. No. 5.6e-58;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

Qy 1 MKKRILSAVLVSGVTLSSTLTSKAVKADDFDAQIASODSKINNLTAAQQAAQAVNTIOG 60
Db 1 MKKRILSAVLVSGVTLSSTLTSKAVKADDFDAQIASODSKINNLTAAQQAAQAVNTIOG 60
Qy 61 QVSALQTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNATSY 120
Db 58 QVSSLSQEQDKLTARNTELEALSQRPEQIKALTSQIVARNEKLNQQAARSAYKNNETSGY 117
Qy 121 INAIINSKVSDAINRVSAIREVVSANERKMLQOEQDKAAVEQKQENQAINTVAANQE 180
Db 118 INALLNSKISDVVNRILVAINRAVSANAKLLEQKADKVSLEEKQANOTAINTIANWA 177
Qy 181 TIAQNTNALNTQQAQLAEALQNLQAEELTTAODKATLVAQKAAAEAEARQAAQAAEA 240
Db 178 MAENQNTLTQQLNVLAATNLAQLASATEKANLVAQKEAEKAAAEALAEQQAQV 237
Qy 241 KAAAEAKALQEAQAAQAAQVAANNNTQATDASDQAAAAADNTQAAQTGDSTEQSAAQAVNNS 300
Db 238 KA-----QEQA-----QQAASVEAAKSAITPAQATPAQSSNAI 273
Qy 301 QDESTTATEAQPSASSASTAAVAANTSSANTYPAGQCTGWKSLAPWVGNWNGGOWAA 360
Db 274 EPAALTA-----PAAPSAGP---QTSYDSSNTYPVGOCTWGAKS LAPWAGNNGGOWAY 326
Qy 361 SAAAAGRVGTSAGAVAVNNDGGYGHVAVYTGVO-GGQIOVQEAANYAGNQSIGNYRGW 419
Db 327 SAQAAGRTGTPWVGAIAVNDGGYGHVAVVVEVQSASSIRVWESNYGRQYIADHRGW 386
Qy 420 FNPQSVSYIPN 431
Db 387 FNPQSVSYIPN 398

RESULT 12
US-10-472-928-4652
; Sequence 4652, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926W
```

```
; CURRENT APPLICATION NUMBER: US/10/472, 928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4652
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: secreted 45 kd protein (usp45)
; OTHER INFORMATION: Cellular location: outside
; OTHER INFORMATION: Similar to strain R6 sequence 15904062 (O.E+01)
US-10-472-928-4652

Query Match 50.3%; Score 1057.5; DB 5; Length 392;
Best Local Similarity 50.8%; Pred. No. 2.7e-56;
Matches 221; Conservative 74; Mismatches 89; Indels 51; Gaps 6;

Qy 1 MKKRILSAVLVSGVTLSSTLTSKAVKADDFDAQIASODSKINNLTAAQQAAQAVNTIOG 60
Db 1 MKKRILSAVLVSGVTLSSTLTSKAVKADDFDAQIASODSKINNLTAAQQAAQAVNTIOG 60
Qy 61 QVSALQTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNATSY 120
Db 61 QVSAIQAEQSNLQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNATSY 120
Qy 121 INAIINSKVSDAINRVSAIREVVSANERKMLQOEQDKAAVEQKQENQAINTVAANQE 180
Db 121 INTIVNSKISTEATISRAVVAAMSEITVSANRKMLEQKADKKAISEKQVANNDAINVTIANQQ 180
Qy 181 TIAQNTNALNTQQAQLAEALQNLQAEELTTAODKATLVAQKAAAEAEARQAAQAAEA 240
Db 181 KLADDAQALTTKQAEALAEALAEALAEALAEALAEALAEALAEALAEALAEALAEALAE 240
Qy 241 KAAAEAKALQEAQAAQAAQVAANNNTQATDASDQAAAAADNTQAAQTGDSTEQSAAQAVNNS 300
Db 241 KRAAQ-----QSVLASANTNLT-----AQVAQVSESA----- 270
Qy 301 QDESTTATEAQPSASSASTAAVAANTSSANTYPAGQCTGWKSLAPWVGNWNGGOWAA 360
Db 271 -----PVRKVRPTYS-----TNASSYPIGECTGWKSLAPWAGDYGNGAQWAT 315
Qy 361 SAAAAGRVGTSAGAVAVNNDGGYGHVAVYTGVO-GGQIOVQEAANYAGNQSIGNYRGW 419
Db 316 SAAAAGRTGTPQVGAIAVNDGGYGHVAVVTAVESTTRIQVSESNYAGNRTIGNHRGW 375
Qy 420 FNP-----GSVSYIY 429
Db 376 FNPETTSSEGFVIYIY 390

RESULT 13
US-10-617-320-3230
; Sequence 3230, Application US/10617320
; Publication No. US2005013640A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER: <Unknown>
```

```
; OPERATING SYSTEM: <unknown>
; SOFTWARE: <unknown>
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/617,320
;   FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/09/107,433
;   FILING DATE: 30-Jun-1998
;   APPLICATION NUMBER: 60/085131
;   FILING DATE: May 12, 1998
;   APPLICATION NUMBER: 60/051553
;   FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Ariniello, Pamela Deneke
;   REGISTRATION NUMBER: 40,489
;   REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (781)893-5007
;   TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3230:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 399 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
;   ORGANISM: Streptococcus pneumoniae
;   FEATURE:
;     NAME/KEY: misc feature
;     LOCATION: (B) LOCATION 1...399
;   SEQUENCE DESCRIPTION: SEQ ID NO: 3230:
US-10-617-320-3230

Query Match      50.3%; Score 1057.5; DB 5; Length 399;
Best Local Similarity 50.8%; Pred. No. 2.8e-56;
Matches 221; Conservative 74; Mismatches 89; Indels 51; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60
DB 8 MKKKILASLLSTVWVSQVAVLTAAHETTTDDKIAQDNKISNLTAQQQAQKQVDQIQE 67
QY 61 QVSALQTQAELOAENORLEAQSATLGGQITQLSSKIVARNESLKQQAARSQAQSNATSY 120
DB 68 QVSAIQAEQSNLQAEENDRLQAESKKLEGEITELSKNIVSRNQSLKQQAARSQAQSNATSY 127
QY 121 INAINSKSVSDAINRVSAIREVVSANEKMLQOEQDQKAAVEQKQENQAAINTVAANO 180
DB 128 INTIVNSKSTELAIRVVAEIVSANNKMLEQKADKKAISEKVANNDAINTVIANQQ 187
QY 181 TIAQNTNALNTQQAELAAQLNLQAEELTTAQDQKATLVAQKAAAEAAQAAQAAAE 240
DB 188 KLADDAQALTTQAEKAAELSLAAEKATAEGEKASILEQKAAAEAAEAAVAEAAAYKE 247
QY 241 KAAAEKALQEQAAQAVANNNTQATDASDQQAADNTQAAQTGDSSTQSAQAQVNN 300
DB 248 KRASQ-----QQSVLASANTLT-----AQVQVSESAAA----- 277
QY 301 DQESTTATEAQSASSASTAAVAANTSSANTYPAGQCTGCVKSLAPVGNVYNGGQWAA 360
DB 278 -----PVRAKVRPTYS-----TNASSYPIGCTGCVKSLAPVGNVYNGGQWAT 322
QY 361 SAAAAGRVGTSPTAGAVAVNNDGGYGHVAVTVGQV-GQIQVEANYAGNQSIGNYRGW 419
DB 323 SAAAAGFRTGSTQVGAIAKCNNDGGYGHVAVTVAVESTTTRIQVSESNYAGNRTIGNRGW 382
QY 420 FNP-----GSVSYIY 429
DB 383 FNPSTTSEGFVTIY 397

RESULT 14
US-10-282-122A-57658

; Sequence 57658, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, Liangsu
;   APPLICANT: Zamudio, Carlos
;   APPLICANT: Malone, Cheryl
;   APPLICANT: Haselbeck, Robert
;   APPLICANT: Ohlsen, Karl
;   APPLICANT: Zyskind, Judith
;   APPLICANT: Wall, Daniel
;   APPLICANT: Trawick, John
;   APPLICANT: Carr, Grant
;   APPLICANT: Yamamoto, Robert
;   APPLICANT: Forsyth, R.
;   APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 57658
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Enterococcus faecium
; US-10-282-122A-57658

Query Match      28.1%; Score 591; DB 4; Length 524;
Best Local Similarity 31.0%; Pred. No. 7.4e-28;
Matches 162; Conservative 90; Mismatches 165; Indels 106; Gaps 14;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60
DB 1 MKKSLISAVVVCSTLTAVASPTAAADDFDSIQOQDQKIADLKNQADAAQSQIDALES 60
QY 61 QVSALQTQAELOAENORLEAQSATLGGQITQLSSKIVARNESLKQQAARSQAQSNATSY 120
DB 61 QVSEINTQADLLAKQDTRLQESQAVLVKDIADLQERIEKREDTIQKQAREAVSNTSSNY 120
QY 121 INAINSKSVSDAINRVSAIREVVSANEKMLQOEQDQKAAVEQKQENQAAINTVAANO 180
DB 121 IDAVLNADSLADIAIGRVQAMTTVMKANNDLMEQKQDQKKAEDAKKELAEAO 180
QY 181 TIAQNTNALNTQQAELAAQLNLQAEELTTAQDQKATLVAQKAAAEAAEAAE 226
DB 181 ALESQKGLLSKQADLNVLKTSLAAEQATDAEDKADLNROKAEAEQARIRQORLAEQ 240
QY 227 AARQAAAAAAAEKAAAEKALQEQAAQ-AQVAANNNTQATDAS-----DQAAAAADNTQ 281
DB 241 ARQQAQAEKAEKAEQAEAEQATQASSTQAQSSATESSATQSSMTTESSATQSSATE 300
QY 282 AQTGDSSTQSAQAQVNNSDQESTTATEA-----QPSASSASTAAVAANTSSANT 331
```

```
Db      301 ESTTPESSTESTAPESSTESTTTPESSTESTTTPESSTESTTTPESSTEST 360
      332 YPA-----GCTWGVKSLAP----- 346
Db      361 TPAPTTPTDQSDVTGNGT-GSSTPAPTPTTPEQPKVTPAPAPSGSVNGAAIVAEAYK 419
      347 WVGNG--YWGNG-----GOWAASAAAAGYRVG-STPSAGAVAV 380
Db      420 YIGTPYVWGKDPGSPGCFTRVYVQVTRDGGTVPQESAGTKISVQAKAGDLLF 479
      381 W-NDGGYGHVAYTVGQGGQIQVQAEANYAGNQSIGNYRGWFPN 422
Db      480 WSGQGGTYHVAIALG--GGQ-YIHAPQPGESVKVGSVQ-WFAP 518
```

```
RESULT 15
US-10-154-251-67
; Sequence 67, Application US/10154251
; Publication No. US20030092024A1
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 06286-060002
; CURRENT APPLICATION NUMBER: US/10/154,251
; CURRENT FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 10/154,251
; PRIOR FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-154-251-67
```

```
Query Match      26.7%; Score 562; DB 4; Length 210;
Best Local Similarity 55.2%; Pred. No. 1.5e-26;
Matches 116; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

QY      1 MKKRILSAVLVSGVTLLSSATTLGAVKADDFDAIQASQDSKINNLTAAQQQAAQAVNTIQ 60
Db      1 MKKILASLLSTVWVSQVAVLTFAHRETTDDKIAQDNKISNLTAAQQQEAQKQVDIQE 60
QY      61 QVSALQTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNAAQTSY 120
Db      61 QVSAIQAEQSNLQAEQNDRLQAESKLEGEITELSKNIVSRNQSLERKQARSQAQTNQAVTSY 120
QY      121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQEQDKAAVEQKQENQAAINTVAANQE 180
Db      121 INTIVNSKSITEAISRVAAQSEIVSANNKMLEQQQKADKKAISEKQVANNDAINTVIANQQ 180
QY      181 TIAQNTNALNTQQAQLEAAQNLQAEALTTA 210
Db      181 KLADDAQALTTKQAEKXAEKLSAAEKATS 210
```

Search completed: February 10, 2006, 23:25:14  
Job time : 254.865 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:15:51 ; Search time 18.6825 Seconds  
(without alignments)  
302.734 Million cell updates/sec

Title: US-10-797-821-29  
Perfect score: 2104  
Sequence: 1 MKKRILSAVLVSGVTLSSAT.....SIGNVRGWNFGSVIYPN 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 1312538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : 1: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB.pbp:\*  
2: /cgn2\_6/prodata/1/pubpaa/US06\_NEW\_PUB.pbp:\*  
3: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB.pbp:\*  
4: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB.pbp:\*  
5: /cgn2\_6/prodata/1/pubpaa/US09\_NEW\_PUB.pbp:\*  
6: /cgn2\_6/prodata/1/pubpaa/US10\_NEW\_PUB.pbp:\*  
7: /cgn2\_6/prodata/1/pubpaa/US11\_NEW\_PUB.pbp:\*  
8: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB.pbp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2096	99.6	431	7	US-11-052-554A-210 Sequence 210, App
2	1091.5	51.9	398	7	US-11-052-554A-252 Sequence 252, App
3	435.5	20.7	211	7	US-11-052-554A-352 Sequence 352, App
4	409.5	19.5	544	7	US-11-052-554A-358 Sequence 358, App
5	233	11.1	257	6	US-10-793-626-3244 Sequence 3244, App
6	227.5	10.8	971	7	US-11-052-554A-3 Sequence 3, Appl1
7	226	10.7	394	7	US-11-052-554A-79 Sequence 79, Appl1
8	220.5	10.5	1236	6	US-10-873-528-109 Sequence 109, App
9	217	10.3	270	6	US-10-485-517-413 Sequence 413, App
10	215	10.2	149	6	US-10-793-626-1682 Sequence 1682, App
11	215	10.2	157	6	US-10-793-626-2870 Sequence 2870, App
12	200	9.5	1562	7	US-11-052-554A-211 Sequence 211, App
13	195.5	9.3	2101	6	US-10-857-780-23 Sequence 23, Appl1
14	194	9.2	1236	6	US-10-793-626-3154 Sequence 3154, App
15	192	9.1	1586	6	US-10-821-234-901 Sequence 901, App
16	191.5	9.1	1448	6	US-10-485-517-212 Sequence 212, App
17	190	9.0	1410	6	US-10-878-556A-136 Sequence 136, App
18	186	8.8	1107	6	US-10-485-517-145 Sequence 145, App
19	185.5	8.8	758	6	US-10-485-517-144 Sequence 144, App
20	182	8.7	1122	6	US-10-467-657-6112 Sequence 6112, App
21	179.5	8.5	1126	6	US-10-485-517-248 Sequence 248, App
22	178.5	8.5	1290	6	US-10-485-517-141 Sequence 141, App
23	176.5	8.4	330	6	US-10-485-517-415 Sequence 415, App
24	174	8.3	3712	7	US-11-019-711-48 Sequence 48, Appl1
25	174	8.3	3712	7	US-11-019-711-51 Sequence 51, Appl1

26	172	8.2	1565	6	US-10-467-657-2704 Sequence 2704, Ap
27	172	8.2	5024	6	US-10-793-626-2964 Sequence 2964, Ap
28	171.5	8.2	834	7	US-11-052-554A-212 Sequence 212, App
29	167.5	8.0	1992	7	US-11-069-834-58 Sequence 58, Appl1
30	167.5	8.0	2000	7	US-11-069-834-56 Sequence 56, Appl1
31	166	7.9	684	6	US-10-793-626-2098 Sequence 2098, Ap
32	165.5	7.9	761	6	US-10-485-517-252 Sequence 252, App
33	165	7.8	860	7	US-11-019-711-59 Sequence 59, Appl1
34	162.5	7.7	612	6	US-10-467-657-3988 Sequence 3988, Ap
35	162.5	7.7	1634	7	US-11-019-711-49 Sequence 49, Appl1
36	162.5	7.7	1640	7	US-11-019-711-8 Sequence 1016, Ap
37	162.5	7.7	3690	6	US-10-995-561-1016 Sequence 1016, Ap
38	162.5	7.7	3714	6	US-10-995-561-1015 Sequence 1015, Ap
39	162.5	7.7	3717	6	US-10-821-234-1076 Sequence 1076, Ap
40	161.5	7.7	706	6	US-10-485-517-146 Sequence 146, App
41	159.5	7.6	716	7	US-11-194-890-20 Sequence 20, Appl1
42	158.5	7.5	1404	6	US-10-878-556A-169 Sequence 169, App
43	157	7.5	365	7	US-11-052-554A-223 Sequence 223, App
44	156	7.4	655	6	US-10-793-626-1052 Sequence 1052, Ap
45	156	7.4	655	6	US-10-793-626-1400 Sequence 1400, Ap

ALIGNMENTS

RESULT 1  
US-11-052-554A-210  
; Sequence 210, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 210  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans UA159  
US-11-052-554A-210

Query Match	99.6%	Score	2096	DB	7	Length	431
Best Local Similarity	99.5%	Pred. No.	6.1e-113				
Matches	429	Conservative	0	Mismatches	2	Indels	0
Gaps	0						
QY	1	MKKRILSAVLVSGVTLSSATTL	SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG	60			
Db	1	MKKRILSAVLVSGVTLSSATTL	SAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG	60			
QY	61	QVSALQTOQAEIQAENQRLEAQSATL	GGQIQITLSSKIVARNESLKKQQAASAKSNAATSY	120			
Db	61	QVSALQTOQAEIQAENQRLEAQSATL	GGQIQITLSSKIVARNESLKKQQAASAKSNAATSY	120			
QY	121	INAINSKSVDAINRVSAIREVVSANEKMLQOEODKAAVEKQOENQAANTVAANO	E	180			
Db	121	INAINSKSVDAINRVSAIREVVSANEKMLQOEODKAAVEKQOENQAANTVAANO	E	180			
QY	181	TTAQNNTALNTQQAOLLEAAQLNLQAEITTTA	QDKATLVAKQAEEAAEAQAAAAQAAAAEA	240			
Db	181	TTAQNNTALNTQQAOLLEAAQLNLQAEITTTA	QDKATLVAKQAEEAAEAQAAAAQAAAAEA	240			
QY	241	KAAAEAKALQEQAAQAQVAANNNTQATDASD	QOAAAAADNTQAAQTGDSSTEQSAQAQVNNNS	300			
Db	241	KAAAEAKALQEQAAQAQVAANNNTQATDASD	QOAAAAADNTQAAQTGDSSTEQSAQAQVNNNS	300			
QY	301	DOESTTATEAQPSASSASTAAVAANTSSANT	TPAGQCTGWGKSLAPWGVNNGGQWAA	360			

```
Db 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGCTWGVKSLAPWVGNWNGGQWAA 360
QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAYVTGVGGQIQVOEANYAGNQSIGNRGWF 420
Db 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAYVTGVGGQIQVOEANYAGNQSIGNRGWF 420
QY 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431

RESULT 2
US-11-052-554A-252
; Sequence 252, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 252
; LENGTH: 398
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes MGAS8232
US-11-052-554A-252

Query Match 51.9%; Score 1091.5; DB 7; Length 398;
Best Local Similarity 53.9%; Pred. No. 9e-56;
Matches 233; Conservative 57; Mismatches 107; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAQVNTTIG 60
Db 1 MKKRILSAVLVSGVTLSAATT---VGAEDLSKIAKQDSIISNLTTEQKAAQNOVSALQA 57

QY 61 QVSALQTOQAEIQAENORLEAQSATLGGQIQIOTLSSKIVARNESLKOQARSNAQTSY 120
Db 58 QVSSLOSEODKLARTTELEALSKEFEQIKALTSQIVARNEKLKQARSAYKNETSQY 117

QY 121 INAINSKSVDAINRVSAREVSVANEKMLQQEQDKAAVEQKQOENQAAINTVAANOE 180
Db 118 INALLNSKISIDWNRVLAINRAVSANAKLLEQKADKVSLEEKQAAQNTAINTIAANWA 177

QY 181 TTAONTNALTQOAELEAQAOLMLQAEFLTQAQOKATLVAQKAAABEAARQAAAAQAABA 240
Db 178 MAEENQNTLRTQOANLEATANLALQLAGATEDKANLVAQKEAAEKAAAEALAEQAAKV 237

QY 241 KAAAEAKALQEAAQAQVAANNNTQATDASDQOAAADNTQAAQTGDSFEQSAQAQVANS 300
Db 238 KA-----QEQA-----QQAASVEAKSAITPAQATPAQGSNAI 273

QY 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGCTWGVKSLAPWVGNWNGGQWAA 360
Db 274 EPAALTA---PAAPSARP---QTSYDSSNTTYPVGGCTWGAKSLAPWAGNNGNGGQW 326

QY 361 SAAAGYRVGSTPSAGAVAVNDGGYGHVAYVTGVQ-GGQIQVOEANYAGNQSIGNRGW 419
Db 327 SAQAAGYRTGSTPMVGAIAVNDGGYGHVAVVVEVQSASSIRVMESNYSGRQYIADHRGW 386

QY 420 NPGSVSYIYPN 431
Db 387 FNPTGVTFLYPH 398

RESULT 3
```

```
US-11-052-554A-352
; Sequence 352, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 352
; LENGTH: 211
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-352

Query Match 20.7%; Score 435.5; DB 7; Length 211;
Best Local Similarity 42.0%; Pred. No. 1e-18;
Matches 92; Conservative 29; Mismatches 77; Indels 21; Gaps 5;

QY 220 OKAAAEERARQAAAAQAABEAKAKALQEAAQAQVAANNNTQATDASDQOAAADN 279
Db 7 EXAVFTVAA--TAAIVVLGNKWKADADTYTLTGDSGDFFSVAQRYHMDAYELASMN--GKDI 62

QY 280 TQAQQTGDSFEQSAQAQVANSQESTTATEAQPSSASTAAVAANTSSANTYPAGQCTW 339
Db 63 TSLILPGQTLTVNGSAAPDNQAAAPTDTTQ-----ATTETNDANANTYPVGCQCTW 112

QY 340 GVKSLAPVGNVWNGGQWAAAGYRVGSTPSAGAVAVNDGGYGHVAYVTGV-QGG 398
Db 113 GKAVATWAGDWNGGQWASSASAGQYTVGNTPAVGSMCTDGGYGHVAVYTVAGEDG 172

QY 399 QIQVOEANYAGNOSIGNYRGWEN-----PGSVSYIYPN 431
Db 173 KVQVLESNYKQQQWVDNRYRGWFDPNNSGDPGVSYSIYPN 211

RESULT 4
US-11-052-554A-358
; Sequence 358, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 358
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-358

Query Match 19.5%; Score 409.5; DB 7; Length 544;
Best Local Similarity 25.2%; Pred. No. 8.9e-17;
Matches 142; Conservative 74; Mismatches 192; Indels 155; Gaps 17;

QY 4 RILSAVLVSGVTLSATTLSAVKADDFDAQIA-----SQDSKINNLTAAQOAAQAQVNTTQ 59
Db 2 RKLKVALFASSILGLMAVSSYTAADTEDNQVTISHYNEQAGTFDVAQVAAANGKTIQSID 61
```



```
Qy 60 GQVSAQTQQAELQ-----AENORLEAQS-----ATL 86
Db 62 VAIWSEENGQDLDKWHASNDGSNQLTVHFAENHGSKVGSYIAHAYITYTGDNRVGVNL 121
Qy 87 GQIQTLSSKIVARNESLQ---QARSQAQNAATSYINAIINSKVSDAINRVSA--- 139
Db 122 GKRKLUSL-----APQLSLKQGLQLFSLKLPKPSADQLFSVWSDENGDDLUHWTADAG 177
Qy 140 -----IREVVSANEKMLQQOEQD-----KAAVE-QKQOENQAAINTVAAN 178
Db 178 NTLAGYANHKGTYVHVHTLKLQNGKMPISQAQDIDIPKPKVIQIDKINDTSYDVVVNN 237
Qy 179 -----OETTAQNTNALNTOQAQLAQLMLQAE-- 206
Db 238 VPPYISSVAIPWSEONGDLDKWKYQATKVADGIFKTTVYLKTRFELGNYQAHYIGDSQ 297
Qy 207 -----LTTAODQKATL-----VAQKAAAEARQAQAAAA 238
Db 298 LSKKLDGLGETHFNVPISINVEYDQVTDIDYHINKGTFDVTVAETDNSKAIQSISAAVWS 357
Qy 239 EAKAA-----AEAKALQEQAAQVAANNTQATDASDQAAAAADNTQAAQTGDSQSA 294
Db 358 DANQANLYWEAKQLANGKRAITVDQKHGQNGTGSYNVHVYHYN-----DGTSGHV 410
Qy 295 QAVNNSDOESTTATEAOPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWGN 354
Db 411 LA-----NQNLQIVHYQPSA-----VRITATYMEKNTYPVGQCTWGVKELAPWIPNMLGN 461
Qy 355 GGOWAASAAAAGYRVGSTPSAGAVAVNDGGYGHVAVYTGVO-QGQIQVOEANYAGNOSI 413
Db 462 GGQWASTAVKFGKIGTVKPVKGALACWSDGGYGHVAVYTHVESNNRIQVKEANYKNOQYI 521
Qy 414 GNYRGWENP-----GSVSYIYPN 431
Db 522 SNFRGWFDPITSYLGRLTYIPD 544

RESULT 5
US-10-793-626-3244
; Sequence 3244, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3244

Query Match 11.1%; Score 233; DB 6; Length 257;
Best Local Similarity 34.5%; Pred. No. 4.3e-07;
Matches 59; Conservative 29; Mismatches 75; Indels 8; Gaps 5;

Qy 261 NNNTQATDASDQAAAAADNTQAAQTGDSQSAQAQAVNNSDOESTTATEAOPSASSASTA 320
Db 91 NNYSNYNNYNNYQ---SNNTSQRTQPTGGLGASYSTSSNVHVTTSA-PSSNGVSL 146
Qy 321 AVAANTSANTYPAGQCTWGV-KSLAPWVGNWGNWGNWGNWGNWGNWGNWGNWGNWGNWGN 379
Db 147 --NARSAGNLYTSGQCTYYVDFRVGKIGSTWGNWGNWGNWGNWGNWGNWGNWGNWGNWGN 204
```

```
Qy 380 VWNDGGYGHVAVYTGTV-QGGQIQVOEANYAGNQSIGNYRGHFNPGSVSYIY 429
Db 205 QTSQGYGHVAVYEGVNSNGSIRVSEMYHGHGAGVVTSTRTISASQAASYN 255

RESULT 6
US-11-052-554A-3
; Sequence 3, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
; LENGTH: 971
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-3

Query Match 10.8%; Score 227.5; DB 7; Length 971;
Best Local Similarity 25.5%; Pred. No. 4e-06;
Matches 92; Conservative 58; Mismatches 172; Indels 39; Gaps 8;

Qy 19 ATTLSAVKADDFDAQIASQDSKINNLTAQOQAAQAVNTIQGVSAQTQQAELQAEQNR 78
Db 121 AQNTAAAKKSADASTSAREAAATHTADADSARAASTAGQAASSAQSSASAGTASTKA 180
Qy 79 LEQASATLGGQIQTLSSKIVARNESLQQAQSAQKSNAAATSYINAIINSKVSDAINRVS 138
Db 181 TEA-----SKSAAAEESSKSAATSETN-ARTSETNAAVVSQQA-----TS 221
Qy 139 AIREVWSANEKMLQQOEQDKAAVEQKQENQAAINTVAANQETIAQNTNALNTQQAQLEA 198
Db 222 ASTATTKASAAASARDASASKEAAKSETSAASSASSA-----ASSATAAGNSAKAAKT 276
Qy 199 AQLNLQALHTTAQDQKATLVAKAAAEBAEAQAAAAQAAAAKAAKALQEQAAQAV 258
Db 277 SETNAKSSETAEEQASAGAAAGSKTAAALSASAASTAGQASASATAAGKSAESAASSAST 336
Qy 259 AANNTOATDASDQAAAAADNTQAAQTGDSQSAQAQAVNNSDOESTTATEAOPSASSAS 318
Db 337 A---TTKAGATEQAASAAASASAASAKTSETN---AKASETSAESSKTA-----AASSAS 384
Qy 319 TAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWGNWGNWGNWGNWGNWGNWGNWGNWGN 378
Db 385 SAASSASSASAKDEATROASAASKSATTAAT---KATEAAGSATAAQAQSKSTAESAATR 441
Qy 379 A 379
Db 442 A 442

RESULT 7
US-11-052-554A-79
; Sequence 79, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
```

```
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 79
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-79

Query Match      10.7%; Score 226; DB 7; Length 394;
Best Local Similarity 27.3%; Pred. No. 1.7e-06;
Matches 108; Conservative 63; Mismatches 177; Indels 48; Gaps 15;

QY 5 ILSNVLGVTLSSATTLSSAVKADDFDAQIASQD-----DSKINNLTAAQQAQAVNTIQ 59
Db 16 IISAVL--HVLFAALWSS-----FDENIEASAGGGSSIDAVMVDSGAVVEQYKRMQ 68
QY 60 GOVSALQTOQAELEQAEORLEAQSATLGGQIQIOTLSSKIVARNESLKO-----QARSAAKS 114
Db 69 SQSSS-----AKSDEQRKKKEQQA-----EELREKQAAEQERLKKOLEKRLAAQEQKK 118
QY 115 NAATSYINAINIKSVSDAINRVSAIREVVSANEKMLQQOQEDKAAVEQKQOENQAAINT 174
Db 119 QAEEAAQAELEKQAEAEAAKAAADAKAKAEADDKAAEEAAKAAADAKKAAEAAKA 178
QY 175 VAANOETIAONTNALNTQOQLEAQLNLQ---AELTTAQDQKATLVAAKAAAEAA--R 229
Db 179 AAEAQKAAEAAAALKKKBAEAEAAAEAAAEAAEKAAAEAAADKKA---AEKAAAEKAAADK 235
QY 230 QAAAAQAAAEAAKAAAEAKALQEOAAQAVANNNTQATDASDQQAADNTQAAQTGDST 289
Db 236 KAAAEKAAADKAAA-AKAAAEKAAAKAAA-----EADDIFEGELSSGKNAPKTGGGA 287
QY 290 EQSAAQAQAVNNSDOESTTATEAQ-PSASSASTAAVAANTSSANTYPAGQCTGWKSLAP-- 346
Db 288 KGNASPAAGSNGTKNKGASGADINNYAGQIKSAIESKFYDASSVAGTKTCLRIK-LAPDG 346
QY 347 WVGNYWNGGQWA-ASAAAAGYRVGSTPAGAVAV 381
Db 347 MLLDIKPEGDPALCOAALAAAKLAKIPKPPSQAVY 382

RESULT 8
US-10-873-528-109
; Sequence 109, Application US/10873528
; Publication No. US20050276814A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129W0
; CURRENT FILING DATE: 2004-06-23
; PRIOR FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 109
; LENGTH: 1236
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-109

Query Match      10.5%; Score 220.5; DB 6; Length 1236;
Best Local Similarity 21.2%; Pred. No. 1.3e-05;
Matches 83; Conservative 90; Mismatches 191; Indels 27; Gaps 7;

; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patentin version 3.3
; SEQ ID NO 79
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-10-797-821-29.rapbn

Query Match      10.3%; Score 217; DB 6; Length 270;
Best Local Similarity 28.4%; Pred. No. 3.7e-06;
Matches 71; Conservative 29; Mismatches 94; Indels 56; Gaps 8;

QY 222 AAEEAARQAAAAQAAAAEAKAAAEAK-----ALQEQAAQAAVANN-- 263
Db 9 APAITTAAGAAAVLSHDDAEASTQHKKVQSGESLWTIAQQVNTSVESIKQNNLNLSNNVFP 68
QY 264 -----TQATDASDQQAADNTQAAQTGDSTEQSA-----AQAVNN-- 299
Db 69 GOVINVGSGASONTSSNTSSASSHTVVA--GESLNIITANKYGVSDALMQANHLNGYL 126
QY 300 --SDQESTTATEAQPSASSASTAAVAANTSS-----ANTYPAGQCTGWG----KSLAPWV 348
Db 127 IMPNQILTIIPNGSGSGSGGTATQTSNGVNTSPSFNHNLYTEGQCTWYVDFDKSQAGKPI 186
QY 349 GNYWNGGQWAASAAAAGYRVGSTPAGAVAVVNDGGYGHVAVTVTGQG--GQTOVEANY 407
```

```
Db 187 STWSDAKYWSNANDGYQVDNTPSVGAIMQSTPGYPGYGHVAYVERINGDGSILISEMNY 246
Qy 408 AGNQSIGNRY 417
Db 247 ANGPYNNMYR 256

RESULT 10
US-10-793-626-1682
; Sequence 1682, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1682
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1682

Query Match 10.2%; Score 215; DB 6; Length 149;
Best Local Similarity 39.7%; Pred. No. 2.4e-06;
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;

Qy 286 GDSTEQSAAQAVNNSDOESTTATEAQPSSASSASTAAVAANTSSANTYPAGCTWGV-KSL 344
Db 20 GHTTHADAAE--NNNQOOST-----YNYSTTEVSP-SNSGNLYTSGCTWYVYDKT 67

Qy 345 APWVGNWNGGOWAASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAYVTGVQG-GQIQVQ 403
Db 68 GKGIGTWGNANSWATAAQAAGFTVNTTPEGAIMQSSEGAFGHVAFVESVNDGSIIVS 127

Qy 404 EANYAG 409
Db 128 EMNYDG 133

RESULT 11
US-10-793-626-2870
; Sequence 2870, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2870
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2870

Query Match 10.2%; Score 215; DB 6; Length 157;
Best Local Similarity 39.7%; Pred. No. 2.6e-06;
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;
```

```
Qy 286 GDSTEQSAAQAVNNSDOESTTATEAQPSSASSASTAAVAANTSSANTYPAGCTWGV-KSL 344
Db 20 GHTTHADAAE--NNNQOOST-----YNYSTTEVSP-SNSGNLYTSGCTWYVYDKT 67

Qy 345 APWVGNWNGGOWAASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAYVTGVQG-GQIQVQ 403
Db 68 GKGIGTWGNANSWATAAQAAGFTVNTTPEGAIMQSSEGAFGHVAFVESVNDGSIIVS 127

Qy 404 EANYAG 409
Db 128 EMNYDG 133

RESULT 12
US-11-052-554A-211
; Sequence 211, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 211
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-211

Query Match 9.5%; Score 200; DB 7; Length 1562;
Best Local Similarity 25.6%; Pred. No. 0.00025;
Matches 84; Conservative 62; Mismatches 114; Indels 68; Gaps 12;

Qy 28 DDFDAQIASODSKINNLTAAQQAAQOVNTTIGQVVSALQTQQAELQAEENORLEAQSAT-- 85
Db 151 DQYKSDVAHEAEVAKIKAKNQATKEQ-----YEKDMAHKAKEVERINANAASK 200

Qy 86 -----LGQIQITLSS--KIVARNESLKQQAOSA-----QKSNA--TSYINAI-- 125
Db 201 TAYEAKLAQYQADLAQVQKTNAANAAQYQKALAAQYQAEKRVQEAANAAKAAYDTAVAA 260

Qy 126 NSKSVSDAINRVSAIREVVSANEKMLQQOQODKAAVEKQOEQAANTVAANOETIAQN 185
Db 261 NAKNT-----EIAAANEIRKRNATAKAETKLAQYQAEKRV-----QE 301

Qy 186 TNAITQQQAQLEAAQLNLQAEELTTAODQKATLVQAQKAAAEAAQAAAAQAAAAEAKAAAE 245
Db 302 ANAAN-----EADYQAKLTAYQTELARVQKANAADAKAAAYEAAVANAANKWAALTAE 352

Qy 246 AKALQEQAAQAVQV---AANNNTQATDASDQQAQAADNT--QAAQTGDTSEQSAQAQVNN 300
Db 353 NTAIKQENENAKATYEALAKQYEAADLAQVKKANANEAADYQAKLTAYQTELARVQKAN-- 410

Qy 301 DQESTTATEAQPSSASSASTAAVAANTSS 328
Db 411 -ADAKAAYEAAVAANAANAALTAENTA 437

RESULT 13
US-10-857-780-23
; Sequence 23, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
```

APPLICANT: KAMMERER, STEFAN M.  
APPLICANT: NELSON, MATTHEW ROBERTS  
APPLICANT: RENELAND, RIKARD HENRY  
APPLICANT: HOYAL-WRIGHTSON, CAROLYN R.  
TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS  
TITLE OF INVENTION: THEOREOF  
FILE REFERENCE: SEQ-4069-CP  
CURRENT APPLICATION NUMBER: US/10/857,780  
CURRENT FILING DATE: 2004-05-28  
CURRENT FILING DATE: 2003-11-25  
PRIOR FILING DATE: 2003-11-25  
PRIOR FILING DATE: 2003-07-24  
PRIOR FILING DATE: 2003-07-24  
PRIOR FILING DATE: 2003-11-25  
NUMBER OF SEQ ID NOS: 4962  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 23  
LENGTH: 2101  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-857-780-23

Query Match 9.3%; Score 195.5; DB 6; Length 2101;  
Best Local Similarity 23.1%; Pred. No. 0.00064;  
Matches 78; Conservative 65; Mismatches 153; Indels 41; Gaps 10;  
QY 2 KXRLSALVSGVTLSSATLISA-----VKADDFAQIASODSKI-----NNLTAQOQAA 51  
Db 405 KGEVLGDLVLETLKQEAATLAANNLTQOLARVEMLETERGQOQAKLAERGHFEFEKQOL 464  
QY 52 QAQVNTIQGVSAALQOQAE-----QAENORLQAQSATLQGOQTOTLSSKIVARNE---SL 104  
Db 465 SSLITLQSSISNLQAKLEQASQHGARLTAQVASLTSELTTLNATTIQQOQDQELAGL 524  
QY 105 KQARSAQKSNATSYINAIINSKVSDAINRYSAIREVVSANEKMLQQOQEQDQAAVEQK 164  
Db 525 KQAKKEKQAQLAQT-----LQQEQASQGLRHQVEQLSSSLKQKEQQLKEVAEK 573  
QY 165 Q-----QENQAAINTVAANTQIAQNTVALNTQOQAE-----AAQIN-LQAEITTAQOQKAT 216  
Db 574 QEATRODHAQOLATAAEEREASLRERDAAALKQLEALEKEKAALKLETLQQOQLQVANEARDS 633  
QY 217 LVAKAAAEFAAQAAQAAQAAAKAAAKALQEQAAQOQVA---ANNNTQATDASDOQ 273  
Db 634 --AQTSTQQRKAEKLSRVEELQACVETARQEQHEAQVAAELQLRSEQOQKATEKE 691  
QY 274 AAAADNTQAAQTGDSTEQSAAQAVNNSDQESTTATEA 310  
Db 692 RVAQEKDQLQEQALQALKESLKVTKGSLSEBKRRRAADA 728

RESULT 14  
US-10-793-626-3154  
; Sequence 3154, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3154  
; LENGTH: 1095  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence

US-10-793-626-3154  
Query Match 9.2%; Score 194; DB 6; Length 1095;  
Best Local Similarity 22.5%; Pred. No. 0.00037;  
Matches 92; Conservative 61; Mismatches 165; Indels 90; Gaps 12;  
QY 13 GVTLSATLTSVAKADDFDAQIASODSKINNLTAAQOQAAQAVNTIQGVSAALQOQAE 72  
Db 340 GMTTDTANNYKSKKREAE-ELQKAAQIINNGDATEQOITNETNRVQAINAKAKNDL 398  
QY 73 QAENORLEQAQSATLQGOQTOTLSSKIVA-RNESLKKQARSQAQKSNATSYINAIINS----- 127  
Db 399 RADKSQLENAYNQLIQNVDTNGKGPASIQOYQAARQAIETQYNNAKSEAHQILENSPSV 458  
QY 128 KSVSDAINEVSALREVVSANEKMLQQOQEQDQAAVEKQOENQA-----AINT 174  
Db 459 NEVAQALQKVEAVQLKVDNDAIHMLQNKENSALVTAKNLOQQAQVNDQPLTTGTQDSINN 518  
QY 175 VAA-----NOETIAQNTNA-----LNTQOQAELEAAQNLQ-----AELTTA 210  
Db 519 YVAKRNEAQSAIRNAEAVINNGDATAKQISDEKSKVEQALAHLNDAKQQLTADTTTELQTA 578  
QY 211 QDQ-----KATLVAQKAAAEAAQAAQAAQAAAEAKAAAEKALQEQAAQOQVA 259  
Db 579 VQQLNRRGDTNNKPRISINAYNKAIQSLQETQITSAKDANAVIQKPIRTVQEVNNALQOV 638  
QY 260 ANNNTQATDASDOQAAADN-----TQAAQTGDSTEQS----- 292  
Db 639 NQLNQQLTEATINQLOPLSNNDALKAARLNLENKINTVQTDGMTQOQSIEAYQNAKRVQN 698  
QY 293 ----AAQAVNNSD---QESTTATEAQPSASSASTAAVAANTSSANTYP 333  
Db 699 ESNLTALINNGDADSQIITETDRVNOQTNLNLTQAI--NGLTVNKEP 744

RESULT 15  
US-10-821-234-901  
; Sequence 901, Application US/10821234  
; Publication No. US20050255114A1  
; GENERAL INFORMATION:  
; APPLICANT: Labat, Ivan  
; APPLICANT: Stache-Crain, Birgit  
; APPLICANT: Andarmani, Susan  
; APPLICANT: Tang, Y. Tom  
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia  
; FILE REFERENCE: 821A  
; CURRENT APPLICATION NUMBER: US/10/821,234  
; CURRENT FILING DATE: 2004-04-07  
; PRIOR APPLICATION NUMBER: US 60/462,047  
; PRIOR FILING DATE: 2003-04-07  
; NUMBER OF SEQ ID NOS: 1704  
; SOFTWARE: pt\_seq\_genes Version 1.0  
; SEQ ID NO 901  
; LENGTH: 1586  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-821-234-901  
Query Match 9.1%; Score 192; DB 6; Length 1586;  
Best Local Similarity 22.3%; Pred. No. 0.00073;  
Matches 86; Conservative 80; Mismatches 121; Indels 98; Gaps 16;  
QY 33 QIASODSKINNLTAAQOQAAQ---AOVNTIQGVSAQO-----TQOAEQAEQRL- 79  
Db 923 QLVAREQETAVQARMQASYRHHVKEVQOQLOKIRTLQEQLENGPNTQLARLQOESILR 982  
QY 80 -----EAQSATLQGOQIQTLSKIVARNESLKQ---QARSQAQKSNATSYINA 123  
Db 983 DALNQATSVQESQNAELAKRLQELSKVSKELVEKSEAVRQDQEQKALEAKAA-AFEKQ 1041  
QY 124 II-----NSKVSDDAINEVSALREV-----VSAANEKMLQQOQEQDQAAVEKQKQ 166  
Db 1042 VLQQLASHRESEALQRLDEVS--RELCHTOSSHASLRADAEKAEQEQOQOQAEHLHSLKQ 1099



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:05:33 ; Search time 311.954 Seconds  
(without alignments)  
607.053 Million cell updates/sec

Title: US-10-797-821-30  
Perfect score: 2104  
Sequence: 1 MKRIISAVLVSGVTLSSAT.....SIGNRGWFNPGSVSYIYPN 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*  
9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2104	100.0	431	7	ADD93650
2	2104	100.0	431	7	AD937273
3	2100	99.8	431	7	ADD93649
4	2100	99.8	431	9	AD937272
5	2100	99.8	431	9	AE931500
6	2096	99.6	431	7	ADD93653
7	2096	99.6	431	9	AD937276
8	2083.5	99.0	432	7	ADD93651
9	2083.5	99.0	432	7	AD937274
10	2074.5	98.6	432	7	ADD93652
11	2074.5	98.6	432	9	AD937275
12	1235	58.7	447	5	ABP29684
13	1235	58.7	447	8	AD98392
14	1235	58.7	447	8	AD98392
15	1235	58.7	447	8	AD981808
16	1235	58.7	447	8	AD979645
17	1092.5	51.9	398	9	AE931542
18	1085.5	51.6	398	5	ABP25919
19	1085.5	51.6	398	8	AD93884
20	1077	51.2	395	5	ABP25918
21	1057.5	50.3	392	6	ABU02747
22	1057.5	50.3	392	8	ADT50226
23	1057.5	50.3	392	8	ADT50165
24	1057.5	50.3	399	8	ADR94595

25	1057.5	50.3	399	9	AEAS8465
26	1056.5	50.2	392	8	ADK47859
27	1056.5	50.2	392	8	ADT50227
28	674	32.0	461	2	AAR14150
29	673	32.0	461	2	AAR14530
30	661.5	31.4	456	5	ABB55584
31	594	28.2	524	6	ABU29734
32	590	28.0	525	7	ADC95468
33	562	26.7	210	2	AA922579
34	528.5	25.1	482	9	ADV16553
35	528.5	25.1	497	7	ADH88105
36	503	23.9	449	2	AA900250
37	503	23.9	449	5	ABP43469
38	503	23.9	449	6	ABU88497
39	503	23.9	449	6	ABU13748
40	503	23.9	449	9	ADV16734
41	503	23.9	449	9	ADY39216
42	503	23.9	450	7	ADH87829
43	454	21.6	422	2	AA900251
44	454	21.6	422	5	ABP43470
45	454	21.6	422	6	ABU88498

ALIGNMENTS

RESULT 1	ADD93650
ID	ADD93650 standard; protein; 431 AA.
XX	ADD93650;
AC	ADD93650;
XX	29-JAN-2004 (first entry)
DT	29-JAN-2004 (first entry)
XX	Streptococcus mutans glucan binding protein-B.
DE	Streptococcus mutans glucan binding protein-B.
XX	Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
KW	Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX	Streptococcus mutans.
OS	Streptococcus mutans.
XX	WO2003075845-A2.
PN	18-SEP-2003.
XX	07-MAR-2003; 2003WO-US006962.
PF	07-MAR-2002; 2002US-0363209P.
PR	08-AUG-2002; 2002US-0402483P.
DR	(FORS-) FORSYTH INST.
XX	Smith DJ, Taubman MA;
PA	WPI; 2003-845091/78.
XX	GENBANK; AY046411.

Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.

Claim 5; Page 8; 49pp; English.

The present sequence is the protein sequence of the glucan binding protein-B (GbpB) of Streptococcus mutans strain 3VF4. The sequence includes a number of human leucocyte antigen (HLA)-binding epitopes. The invention provides immunogenic compositions and vaccines for dental caries. The compositions comprise major histocompatibility complex (MHC) class II protein-binding GbpB peptides, especially HLA-binding peptides, covalently linked with peptide subunits of a glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diepitopic or multi-epitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in

CC	passive immunisation.	
XX		
SQ	Sequence 431 AA;	
	Query Match 100.0%; Score 2104; DB 7; Length 431;	
	Best Local Similarity 100.0%; Pred. No. 3e-134;	
	Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MKKRILSAVLVSGVTLSSTLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60	
DB	1 MKKRILSAVLVSGVTLSSTLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60	
QY	61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSKIVARNESLKQQAQAQVNTIQ 120	
DB	61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSKIVARNESLKQQAQAQVNTIQ 120	
QY	121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQEQDKAAVEKQQAQVNTIAANOE 180	
DB	121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQEQDKAAVEKQQAQVNTIAANOE 180	
QY	181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAQKAAAEBAARQAAAAA 240	
DB	181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAQKAAAEBAARQAAAAA 240	
QY	241 KAAAEKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSFQSAQAQVNS 300	
DB	241 KAAAEKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSFQSAQAQVNS 300	
QY	301 DQESTTATEAQAQPSASASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360	
DB	301 DQESTTATEAQAQPSASASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360	
QY	361 SAAAAGYRVGSTPSAGAVAVMNDGGYGHVAYVTGVGGQIQVQEAANYAGNQSIGNYRGWF 420	
DB	361 SAAAAGYRVGSTPSAGAVAVMNDGGYGHVAYVTGVGGQIQVQEAANYAGNQSIGNYRGWF 420	
QY	421 NPGSVSYIYPN 431	
DB	421 NPGSVSYIYPN 431	
RESULT 2		
AD37273		
ID	AD37273 standard; protein; 431 AA.	
XX		
AC	AD37273;	
XX		
DT	21-APR-2005 (first entry)	
XX		
DE	Streptococcus mutant glucan binding protein B variant #2.	
XX		
KW	immunogenicity; immune stimulation; glucan binding protein-B;	
KW	microparticle; major histocompatibility complex; tooth disease.	
XX		
OS	Streptococcus mutans.	
XX		
PN	US2005031633-A1.	
XX		
PD	10-FEB-2005.	
XX		
PF	09-MAR-2004; 2004US-00797821.	
XX		
PR	13-APR-1998; 98US-0081550P.	
PR	08-JAN-1999; 99US-0115142P.	
PR	12-APR-1999; 99US-00290049.	
PR	07-MAR-2002; 2002US-0363209P.	
PR	08-AUG-2002; 2002US-0402483P.	
PR	07-MAR-2003; 2003US-00383930.	
XX		
PA	(SMIT/) SMITH D J.	
PA	(TAUB/) TAUBMAN M A.	
XX		
PI	Smith DJ, Taubman MA;	

XX	WPI; 2005-151644/16.	
XX		
PT	New composition comprising a fragment of a glucan binding protein-B	
PT	(GbpB) that binds to MHC class II protein, and a biocompatible	
PT	microparticle, useful for producing an antibody (claimed) for immunizing	
PT	mammals against dental caries.	
XX		
PS	Claim 3; SEQ ID NO 30; 73pp; English.	
XX		
CC	The invention relates to a composition comprising a fragment of a glucan	
CC	binding protein-B (GbpB) and a biocompatible microparticle, where the	
CC	fragment binds to a major histocompatibility complex (MHC) class II	
CC	protein. The composition is useful for producing an antibody for	
CC	immunizing mammals against dental caries. This sequence corresponds to a	
CC	Streptococcus mutans GbpB protein of the invention.	
XX		
SQ	Sequence 431 AA;	
	Query Match 100.0%; Score 2104; DB 9; Length 431;	
	Best Local Similarity 100.0%; Pred. No. 3e-134;	
	Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MKKRILSAVLVSGVTLSSTLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60	
DB	1 MKKRILSAVLVSGVTLSSTLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60	
QY	61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSKIVARNESLKQQAQAQVNTIQ 120	
DB	61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSKIVARNESLKQQAQAQVNTIQ 120	
QY	121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQEQDKAAVEKQQAQVNTIAANOE 180	
DB	121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQEQDKAAVEKQQAQVNTIAANOE 180	
QY	181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAQKAAAEBAARQAAAAA 240	
DB	181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAQKAAAEBAARQAAAAA 240	
QY	241 KAAAEKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSFQSAQAQVNS 300	
DB	241 KAAAEKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSFQSAQAQVNS 300	
QY	301 DQESTTATEAQAQPSASASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360	
DB	301 DQESTTATEAQAQPSASASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360	
QY	361 SAAAAGYRVGSTPSAGAVAVMNDGGYGHVAYVTGVGGQIQVQEAANYAGNQSIGNYRGWF 420	
DB	361 SAAAAGYRVGSTPSAGAVAVMNDGGYGHVAYVTGVGGQIQVQEAANYAGNQSIGNYRGWF 420	
QY	421 NPGSVSYIYPN 431	
DB	421 NPGSVSYIYPN 431	
RESULT 3		
AD373649		
ID	AD373649 standard; protein; 431 AA.	
XX		
AC	AD373649;	
XX		
DT	29-JAN-2004 (first entry)	
XX		
DE	Streptococcus mutans glucan binding protein-B.	
XX		
KW	Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.	
XX		
OS	Streptococcus mutans.	
XX		
PH	Key Location/Qualifiers	
FT	Region 6..25	
FT	/note= "HLA-binding peptide"	





```
PI Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
XX
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX Claim 3; SEQ ID NO 29; 73pp; English.
XX
CC The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
XX Sequence 431 AA;
SQ
Query Match 99.8%; Score 2100; DB 9; Length 431;
Best Local Similarity 99.8%; Pred. No. 5.5e-134;
Matches 430; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MKKRILSAVLVSGVTLSATTLSSATTLSSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
Db 1 MKKRILSAVLVSGVTLSATTLSSATTLSSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQQTTLSSKIVARNESLKQQAARSQAQKSNATSY 120
Db 61 QVSALOTQQAELQAEORLEAQSATLGGQQTTLSSKIVARNESLKQQAARSQAQKSNATSY 120
Qy 121 INAIINSKVSVDALNRVSAIREVVSAANEKMLQQEQDKAAVEQKQENQAINTVAANQE 180
Db 121 INAIINSKVSVDALNRVSAIREVVSAANEKMLQQEQDKAAVEQKQENQAINTVAANQE 180
Qy 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAQKAAAEAAARQAAAAA 240
Db 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAQKAAAEAAARQAAAAA 240
Qy 241 KAAAEAKALOEQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTQSAAQAVNNS 300
Db 241 KAAAEAKALOEQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTQSAAQAVNNS 300
Qy 301 DOESTTATEAQPSSASTAAVAANTSANTYPAGQCTWGVKSLAPWGVNNGGQWAA 360
Db 301 DOESTTATEAQPSSASTAAVAANTSANTYPAGQCTWGVKSLAPWGVNNGGQWAA 360
Qy 361 SAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYVTGQGGQIQVQEAANYAGNQSIGNYRGWF 420
Db 361 SAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYVTGQGGQIQVQEAANYAGNQSIGNYRGWF 420
Qy 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431
RESULT 5
AEB91500
ID AEB91500 standard; protein; 431 AA.
XX
AC AEB91500;
XX
DT 20-OCT-2005 (first entry)
XX
DE Microbial pathogen adhesin protein sequence, SEQ ID NO:210.
XX
KW algorithm; adhesin; pharmaceutical; vaccine; drug screening;
KW bordetella pertussis infection; antibacterial; pneumonia;
KW antiinflammatory; respiratory-gen.; gastric ulcer; antiulcer;
KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.
XX
OS Streptococcus mutans.
```

```
XX WO2005076010-A2.
XX 18-AUG-2005.
XX
XX 07-FEB-2005; 2005WO-IN000037.
XX
XX 06-FEB-2004; 2004IN-DE000173.
XX
XX 20-JUL-2004; 2004US-0589227P.
XX
XX (COUL ) COUNCIL SCI & IND RES SOUTH AFRICA.
XX
XX Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramachandran S;
XX WPI; 2005-597835/61.
XX
XX Computational method for identifying adhesin and adhesin like molecules,
XX comprises computing sequence-based attributes of protein sequences using
XX neural network software and training an artificial neural network.
XX
XX Claim 16; SEQ ID NO 210; 402pp; English.
XX
XX The present invention relates to a computational method (M1) for
XX identifying adhesin and adhesin-like proteins, by computing the sequence-
XX based attributes of protein sequences using five attribute modules of a
XX neural network software, training an artificial neural network (ANN) for
XX each of the computed five attributes, and identifying the adhesin and
XX adhesin-like proteins having probability of being an adhesin (Pad) as
XX equal or greater than 0.51. Also claimed is a set of 274 annotated genes
XX encoding adhesin and adhesin-like proteins, having 274 fully defined 162-
XX 1151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical
XX genes encoding adhesin and adhesin-like proteins, having 105 fully
XX defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated
XX adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base
XX pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-
XX like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO:
XX 280-384) sequences; and a fully connected multilayer feed forward ANN (I)
XX based on (M1). (M1) is useful for identifying adhesin and adhesin-like
XX proteins, of therapeutic potential, and identifying and short-listing
XX proteins for further testing in development of new vaccine formulations
XX to eliminate diseases caused by various pathogenic organisms. (M1) is
XX useful for identifying putative adhesins that are important in drug
XX discovery and preventing therapeutics for whooping cough, pneumonia,
XX gastric ulcer and urinary tract infections. (M1) identifies adhesins from
XX distantly related organisms, and from bacteria belonging to a wide
XX phylogenetic spectrum. (M1) is capable of predicting adhesive nature of
XX unique proteins. The present sequence is a microbial pathogen adhesin
XX protein sequence.
XX
XX Sequence 431 AA;
Query Match 99.8%; Score 2100; DB 9; Length 431;
Best Local Similarity 99.8%; Pred. No. 5.5e-134;
Matches 430; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MKKRILSAVLVSGVTLSATTLSSATTLSSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
Db 1 MKKRILSAVLVSGVTLSATTLSSATTLSSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQQTTLSSKIVARNESLKQQAARSQAQKSNATSY 120
Db 61 QVSALOTQQAELQAEORLEAQSATLGGQQTTLSSKIVARNESLKQQAARSQAQKSNATSY 120
Qy 121 INAIINSKVSVDALNRVSAIREVVSAANEKMLQQEQDKAAVEQKQENQAINTVAANQE 180
Db 121 INAIINSKVSVDALNRVSAIREVVSAANEKMLQQEQDKAAVEQKQENQAINTVAANQE 180
Qy 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAQKAAAEAAARQAAAAA 240
Db 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQDQKATLVAQKAAAEAAARQAAAAA 240
Qy 241 KAAAEAKALOEQAQAQAANNNTQATDASDQQAADNTQAAQTGSDTQSAAQAVNNS 300
XX
```

Db 241 KAAAEAKALQEAQAQAANNNTQATDVSDQAAAADNTQAAQTGDSSTESAAQAVNNS 300  
 Qy 301 DQESTTATEAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWAA 360  
 Db 301 DQESTTATEAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWAA 360  
 Qy 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGWF 420  
 Db 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGWF 420  
 Qy 421 NPGSVSYIYPN 431  
 Db 421 NPGSVSYIYPN 431

RESULT 6  
 ADD93653  
 ID ADD93653 standard; protein; 431 AA.  
 XX  
 AC ADD93653;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Streptococcus mutans glucan binding protein-B.  
 XX  
 KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PN WO2003075845-A2.  
 XX  
 PD 18-SEP-2003.  
 XX  
 PF 07-MAR-2003; 2003WO-US006962.  
 XX  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 XX  
 PA (FORS-) FORSYTH INST.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX  
 WPI; 2003-845091/78.  
 DR GENBANK; AY046414.  
 XX  
 PT Composition useful as vaccines for dental caries comprises a fragment of  
 a glucan binding protein-B binding to a major histocompatibility complex  
 class II protein.  
 XX  
 PS Claim 5; Page 8-9; 49pp; English.  
 XX  
 CC The present sequence is the protein sequence of the glucan binding  
 protein-B (GbpB) of Streptococcus mutans strain 5SM3. The sequence  
 includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
 invention provides immunogenic compositions and vaccines for dental  
 caries. The compositions comprise major histocompatibility complex (MHC)  
 class II protein-binding GbpB peptides, especially HLA-binding peptides,  
 covalently linked with peptide subunits of a glucosyltransferase. The  
 compositions are used in a claimed method of eliciting production of an  
 antibody in a mammal. Diepitopic or multiepitopic polypeptides can be  
 prepared synthetically or by recombinant DNA technology. Antibodies  
 raised against MHC class II binding fragments of GbpB can be used in  
 passive immunisation.  
 XX  
 SQ Sequence 431 AA;  
 Query Match 99.6%; Score 2096; DB 7; Length 431;  
 Best Local Similarity 99.5%; Pred. No. 1e-133;  
 Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 MKKRILSAVLVSGVTLSSATTL SAVKADDPDAQIASQDSKINNLTAQOQAAQAVNTIQG 60  
 Db 1 MKKRILSAVLVSGVTLSSATTL SAVKADDPDAQIASQDSKINNLTAQOQAAQAVNTIQG 60

Qy 61 QVSALQTQAELOAENORLEAQSATLGQIQTLSSKIVARNESLKQQAARSQKSNAAATSY 120  
 Db 61 QVSALQTQAELOAENORLEAQSATLGQIQTLSSKIVARNESLKQQAARSQKSNAAATSY 120  
 Qy 121 INAIINSKVSDAINRVSAIREVVVSANEKMLQQQEQDQKAAVEQKQOENQAAINTVAANQE 180  
 Db 121 INAIINSKVSDAINRVSAIREVVVSANEKMLQQQEQDQKAAVEQKQOENQAAINTVAANQE 180  
 Qy 181 TIAQNTNALNTQQAQLEAAQNLQAEELTTAQDQKATLVAQKAAAEAAARQAAAAQAAAEA 240  
 Db 181 TIAQNTNALNTQQAQLEAAQNLQAEELTTAQDQKATLVAQKAAAEAAARQAAAAQAAAEA 240  
 Qy 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSSTESAAQAVNNS 300  
 Db 241 KAAAEAKALQEAQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSSTESAAQAVNNS 300  
 Qy 301 DQESTTATEAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWAA 360  
 Db 301 DQESTTATEAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWAA 360  
 Qy 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGWF 420  
 Db 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVVTGVGGQIQVQEAANYAGNOSIGNYRGWF 420  
 Qy 421 NPGSVSYIYPN 431  
 Db 421 NPGSVSYIYPN 431

RESULT 7  
 ADX37276  
 ID ADX37276 standard; protein; 431 AA.  
 XX  
 AC ADX37276;  
 XX  
 DT 21-APR-2005 (first entry)  
 XX  
 DE Streptococcus mutant glucan binding protein B variant #5.  
 XX  
 KW immunogenicity; immune stimulation; glucan binding protein-B;  
 KW microparticle; major histocompatibility complex; tooth disease.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PN US2005031633-A1.  
 XX  
 PD 10-FEB-2005.  
 XX  
 PF 09-MAR-2004; 2004US-00797821.  
 XX  
 PR 13-APR-1998; 98US-0081550P.  
 PR 08-JAN-1999; 99US-0115142P.  
 PR 12-APR-1999; 99US-002900049.  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 PR 07-MAR-2003; 2003US-00383930.  
 XX  
 PA (SMIT/) SMITH D J.  
 PA (TAUB/) TAUBMAN M A.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX  
 WPI; 2005-151644/16.  
 XX  
 PT New composition comprising a fragment of a glucan binding protein-B  
 (GbpB) that binds to MHC class II protein, and a biocompatible  
 microparticle, useful for producing an antibody (claimed) for immunizing  
 mammals against dental caries.  
 XX  
 PS Claim 3; SEQ ID NO 33; 73pp; English.  
 XX  
 CC The invention relates to a composition comprising a fragment of a glucan

CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II  
 CC protein. The composition is useful for producing an antibody for  
 CC immunizing mammals against dental caries. This sequence corresponds to a  
 CC Streptococcus mutans GbpB protein of the invention.

XX  
 SQ Sequence 431 AA;

Query Match 99.6%; Score 2096; DB 9; Length 431;  
 Best Local Similarity 99.5%; Pred. No. 1e-133;  
 Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MKKRLSAVLVSGVTLSATLSSATLTSVAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60  
 Db 1 MKKRLSAVLVSGVTLSATLSSATLTSVAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60

Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120  
 Db 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120

Qy 121 INAIINSKVSDAINRVSAIREVVSANESKMLQQEQEDKAAVEQKQENQAAINTVAANOE 180  
 Db 121 INAIINSKVSDAINRVSAIREVVSANESKMLQQEQEDKAAVEQKQENQAAINTVAANOE 180

Qy 181 TIAQNTNALNTQQAELQAEALNLQAEILTAAQDOKATILVAQKAAAEAEARQAAAAA 240  
 Db 181 TIAQNTNALNTQQAELQAEALNLQAEILTAAQDOKATILVAQKAAAEAEARQAAAAA 240

Qy 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGSDTEGSAQAQVNS 300  
 Db 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGSDTEGSAQAQVNS 300

Qy 301 DOESTTATEAOPSASSASTAAVAANTSSANTYPAGCCTGWGKSLAPWVGNWNGGOWAA 360  
 Db 301 DOESTTATEAOPSASSASTAVVTANTSSANTYPAGCCTGWGKSLAPWVGNWNGGOWAA 360

Qy 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVVTGQGGQIQVQANVAGNQSIGNYRGWF 420  
 Db 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVVTGQGGQIQVQANVAGNQSIGNYRGWF 420

Qy 421 NPGSVSYIYPN 431  
 Db 421 NPGSVSYIYPN 431

# RESULT 8

ADD93651  
 ID ADD93651 standard; protein; 432 AA.

XX AC ADD93651;

DT 29-JAN-2004 (first entry)

XX DE Streptococcus mutans glucan binding protein-B.

XX KW Glucan binding protein-B; GbpB; vaccine; anticonaries; epitope; immunogen.

XX OS Streptococcus mutans.

XX PN WO2003075845-A2.

XX PD 18-SEP-2003.

XX PF 07-MAR-2003; 2003WO-US006962.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX PA (FORS-) FORSYTH INST.

XX PI Smith DJ, Taubman MA;

XX KW WPI; 2003-845091/78.

DR GENBANK; AY046412.

XX Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 XX class II protein.

XX Claim 5; Page 8; 49pp; English.

XX The present sequence is the protein sequence of the glucan binding  
 CC protein-B (GbpB) of Streptococcus mutans strain 15JP2. The sequence  
 CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
 CC invention provides immunogenic compositions and vaccines for dental  
 CC caries. The compositions comprise major histocompatibility complex (MHC)  
 CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
 CC covalently linked with peptide subunits of a glucosyltransferase. The  
 CC compositions are used in a claimed method of eliciting production of an  
 CC antibody in a mammal. Dipetopic or multipetopic polypeptides can be  
 CC prepared synthetically or by recombinant DNA technology. Antibodies  
 CC raised against MHC class II binding fragments of GbpB can be used in  
 CC passive immunisation.

XX SQ Sequence 432 AA;

Query Match 99.0%; Score 2083.5; DB 7; Length 432;

Best Local Similarity 99.1%; Pred. No. 7.3e-133;

Matches 428; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Qy 1 MKKRLSAVLVSGVTLSATLSSATLTSVAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60

Db 1 MKKRLSAVLVSGVTLSATLSSATLTSVAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60

Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120

Db 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQARSQAQKSNAAATSY 120

Qy 121 INAIINSKVSDAINRVSAIREVVSANESKMLQQEQEDKAAVEQKQENQAAINTVAANOE 180

Db 121 INAIINSKVSDAINRVSAIREVVSANESKMLQQEQEDKAAVEQKQENQAAINTVAANOE 180

Qy 181 TIAQNTNALNTQQAELQAEALNLQAEILTAAQDOKATILVAQKAAAEAEARQAAAAA 240

Db 181 TIAQNTNALNTQQAELQAEALNLQAEILTAAQDOKATILVAQKAAAEAEARQAAAAA 240

Qy 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGSDTEGSAQAQVNN 299

Db 241 KAAAEAKALQEAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGSDTEGSAQAQVNN 300

Qy 300 SDQESTTATEAOPSASSASTAAVAANTSSANTYPAGCCTGWGKSLAPWVGNWNGGOWA 359

Db 301 SDQESTTATEAOPSASSASTAAVAANTSSANTYPAGCCTGWGKSLAPWVGNWNGGOWA 360

Qy 360 ASAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVVTGQGGQIQVQANVAGNQSIGNYRGW 419

Db 361 ASAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVVTGQGGQIQVQANVAGNQSIGNYRGW 420

Qy 420 FNEGSVSYIYPN 431

Db 421 FNEGSVSYIYPN 432

# RESULT 9

ADX37274

ID ADX37274 standard; protein; 432 AA.

XX AC ADX37274;

XX DT 21-APR-2005 (first entry)

XX DE Streptococcus mutant glucan binding protein B variant #3.

XX KW immunogenicity; immune stimulation; glucan binding protein-B;

XX KW microparticle; major histocompatibility complex; tooth disease.

```

OS Streptococcus mutans.
XX US2005031633-A1.
XX
XX PD 10-FEB-2005.
XX
XX PF 09-MAR-2004; 2004US-00797821.
XX
XX PR 13-APR-1998; 98US-0081550P.
XX PR 08-JAN-1999; 99US-0115142P.
XX PR 12-APR-1999; 99US-00290049.
XX PR 07-MAR-2002; 2002US-0363209P.
XX PR 08-AUG-2002; 2002US-0402483P.
XX PR 07-MAR-2003; 2003US-00383930.
XX
XX (SMIT/) SMITH D J.
XX PA (TAUB/) TAUBMAN M A.
XX
XX PI Smith DJ, Taubman MA;
XX
XX DR WPI; 2005-151644/16.
XX
XX PT New composition comprising a fragment of a glucan binding protein-B
XX PT (GbpB) that binds to MHC class II protein, and a biocompatible
XX PT microparticle, useful for producing an antibody (claimed) for immunizing
XX PT mammals against dental caries.
XX
XX PS Claim 3; SEQ ID NO 31; 73pp; English.
XX
XX CC The invention relates to a composition comprising a fragment of a glucan
XX CC binding protein-B (GbpB) and a biocompatible microparticle, where the
XX CC fragment binds to a major histocompatibility complex (MHC) class II
XX CC protein. The composition is useful for producing an antibody for
XX CC immunizing mammals against dental caries. This sequence corresponds to a
XX CC Streptococcus mutans GbpB protein of the invention.
XX
XX SQ Sequence 432 AA;

Query Match 99.0%; Score 2083.5; DB 9; Length 432;
Best Local Similarity 99.1%; Pred. No. 7.3e-133;
Matches 428; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Qy 1 MKKRILSAVLVSGVTLSATLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 1 MKKRILSAVLVSGVTLSATLSAIAKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQARSAQKSNAAATSY 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQARSAQKSNAAATSY 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 121 INAIINSKVSDAINRVSAREVVSANEKMLQQEQEDKAAVEQKQENQAINTVAANQE 180
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 181 TIAQNTNALNTQQAELQAEALNLQAEITTAQDQKATLVAOKAAAEAEARQAAAAA 240
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 181 TIAQNTNALNTQQAELQAEALNLQAEITTAQDQKATLVAOKAAAEAEARQAAAAA 240
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 241 KAAAEAKALQEAQAQAAA-NNNTOATDASDQQAADNTORAOOTGDSTEOSAQAQVNN 299
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 241 KAAAEAKALQEAQAQAAAANNNTQATDASDQQAADNTQAAQTGDSDDQSAQAQVNN 300
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 300 SDESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 359
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 301 SDESTTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 360
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 360 ASAAAAGRYGVSTPSAGAVVNDGGYGHVAYVTGGGQIQVEANYAGNQSIGNYRGW 419
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 361 ASAAAAGRYGVSTPSAGAVVNDGGYGHVAYVTGGGQIQVEANYAGNQSIGNYRGW 420
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 420 FNPGSVSIYIPN 431
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 421 FNPGSVSIYIPN 432
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

RESULT 10
ADD93652
ID ADD93652 standard; protein; 432 AA.
XX
XX AC ADD93652;
XX
XX DT 29-JAN-2004 (first entry)
XX
XX DE Streptococcus mutans glucan binding protein-B.
XX
XX KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
XX OS Streptococcus mutans.
XX
XX PN WO2003075845-A2.
XX
XX PD 18-SEP-2003.
XX
XX PR 07-MAR-2003; 2003WO-US006962.
XX
XX PR 07-MAR-2002; 2002US-0363209P.
XX
XX PR 08-AUG-2002; 2002US-0402483P.
XX
XX PA (FORS-) FORSYTH INST.
XX
XX PI Smith DJ, Taubman MA;
XX
XX DR WPI; 2003-845091/78.
XX
XX DR GENBANK; AY046413.
XX
XX PT Composition useful as vaccines for dental caries comprises a fragment of
XX PT a glucan binding protein-B binding to a major histocompatibility complex
XX PT class II protein.
XX
XX PS Claim 5; Page 8; 49pp; English.
XX
XX CC The present sequence is the protein sequence of the glucan binding
XX CC protein-B (GbpB) of Streptococcus mutans strain 3SN1. The sequence
XX CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The
XX CC invention provides immunogenic compositions and vaccines for dental
XX CC caries. The compositions comprise major histocompatibility complex (MHC)
XX CC class II protein-binding GbpB peptides, especially HLA-binding peptides,
XX CC covalently linked with peptide subunits of a glucosyltransferase. The
XX CC compositions are used in a claimed method of eliciting production of an
XX CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be
XX CC prepared synthetically or by recombinant DNA technology. Antibodies
XX CC raised against MHC class II binding fragments of GbpB can be used in
XX CC passive immunisation.
XX
XX SQ Sequence 432 AA;

Query Match 98.6%; Score 2074.5; DB 7; Length 432;
Best Local Similarity 98.8%; Pred. No. 3e-132;
Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 MKKRILSAVLVSGVTLSATLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 1 MKKRILSAVLVSGVTLSATLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQARSAQKSNAAATSY 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQARSAQKSNAAATSY 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 121 INAIINSKVSDAINRVSAREVVSANEKMLQQEQEDKAAVEQKQENQAINTVAANOE 180
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 121 INAIINSKVSDAINRVSAREVVSANEKMLHQEQEDKAAVEQKQENQAINTVAANOE 180
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 181 TIAQNTNALNTQQAELQAEALNLQAEITTAQDQKATLVAOKAAAEAEARQAAAAA 240
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Qy 181 TIAQNTNALNTQQAELQAEALNLQAEITTAQDQKATLVAOKAAAEAEARQAAAAA 240
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```
Qy 241 KAAAEAKALOEQAQAQAAAA-NNNTQATDASDQQAADNTQAAQTGDTSTQSAQAQAVNN 299
Db 241 KAAAEAKALOEQAQAQAAAA-NNNTQATDASDQQAADNTQAAQTGDTSTQSAQAQAVNN 300
Qy 300 SDQSTTTATEAOPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 359
Db 301 SDQSTTTATAAOPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 360
Qy 360 ASAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTVGVGGQIQVQEANYAGNQSIGNYRGW 419
Db 361 ASAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTVGVGGQIQVQEANYAGNQSIGNYRGW 420
Qy 420 FNPGSVSYIYPN 431
Db 421 FNPGSVSYIYPN 432

RESULT 11
ID ADX37275 standard; protein; 432 AA.
XX
AC ADX37275;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B variant #4.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
XX (SMIT/) SMITH D J.
XX (TAUB/) TAUBMAN M A.
XX
XX Smith DJ, Taubman MA;
XX
XX WPI; 2005-151644/16.
XX
XX New composition comprising a fragment of a glucan binding protein-B
XX (GbpB) that binds to MHC class II protein, and a biocompatible
XX microparticle, useful for producing an antibody (claimed) for immunizing
XX mammals against dental caries.
XX
XX Claim 3; SEQ ID NO 32; 73pp; English.
XX
XX The invention relates to a composition comprising a fragment of a glucan
XX binding protein-B (GbpB) and a biocompatible microparticle, where the
XX fragment binds to a major histocompatibility complex (MHC) class II
XX protein. The composition is useful for producing an antibody for
XX immunizing mammals against dental caries. This sequence corresponds to a
XX Streptococcus mutans GbpB protein of the invention.
XX
XX Sequence 432 AA;
XX
XX Query Match 98.6%; Score 2074.5; DB 9; Length 432;
XX Best Local Similarity 98.8%; Pred. No. 3e-132;
XX Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;
XX
XX 1 MKKRILSAVLVSGVTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
```

```
Db 1 MKKRILSAVLVSGVTLSAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQG 60
Qy 61 QVSALQTQQAELQAEQENRLEAQAQATLGGQIQTLSSKIVARNESLKQAARSAQKSNAAATSY 120
Db 61 QVSALQTQQAELQAEQENRLEAQAQATLGGQIQTLSSKIVARNESLKQAARSAQKSNAAATSY 120
Qy 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEOKQEQENQAANTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLHQEQDKAAVEQKHQENQAANTVAANQE 180
Qy 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAQDQKATLVAQKAAAEAAARQAAAAQAAAEA 240
Db 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAQDQKATLVAQKAAAEAAARQAAAAQAAAEA 240
Qy 241 KAAAEAKALOEQAQAQAAAA-NNNTQATDASDQQAADNTQAAQTGDTSTQSAQAQAVNN 299
Db 241 KAAAEAKALOEQAQAQAAAA-NNNTQATDASDQQAADNTQAAQTGDTSTQSAQAQAVNN 300
Qy 300 SDQSTTTATEAOPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 359
Db 301 SDQSTTTATAAOPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWA 360
Qy 360 ASAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTVGVGGQIQVQEANYAGNQSIGNYRGW 419
Db 361 ASAAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYTVGVGGQIQVQEANYAGNQSIGNYRGW 420
Qy 420 FNPGSVSYIYPN 431
Db 421 FNPGSVSYIYPN 432

RESULT 12
ABP29684
ID ABP29684 standard; protein; 447 AA.
XX
AC ABP29684;
XX
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polypeptide SEQ ID NO 8544.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS Streptococcus agalactiae.
XX
XX WO200234771-A2.
XX
XX 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB004789.
XX
XX 27-OCT-2000; 2000GB-00026333.
XX
XX 24-NOV-2000; 2000GB-00028727.
XX
XX 07-MAR-2001; 2001GB-00005640.
XX
XX (CHIR-) CHIRON SPA.
XX (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Y RosI, Grandi G, Fraser C;
XX Tettelin H;
XX WPI; 2002-352536/38.
XX DR N-PSDB; ABN70315.
XX
XX New Streptococcus protein for the treatment or prevention of infection or
XX disease caused by Streptococcus bacteria, such as meningitis, and for
XX detecting a compound that binds to the protein.
XX
XX Claim 1; Page 3965; 4525pp; English.
XX
```

CC The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX  
SQ Sequence 447 AA;

Query Match 58.7%; Score 1235; DB 5; Length 447;  
Best Local Similarity 60.1%; Pred. No. 2.4e-75;  
Matches 277; Conservative 43; Mismatches 97; Indels 44; Gaps 9;  
Qy 1 MKKRILSAVLVSGVTLSSATTLSSAVKADDFPAQIASQDSKINNLTAAQQAAQAQVNTIQ 60  
Db 1 MKKRILSAVLVSGVTLGTAAV--TVNADDFDSKIAATDSVINTLSGGQAAQNVTAIKG 58  
Qy 61 QVSLQTOQAQLEAENORLEAQSATLGQOIQTLSSKIVARNESLKQQAARSQAKSNAATSY 120  
Db 59 QVGALESQQSELEAQAQLEAVSQOLGOEIQTLNKNIVARNESLKQVRSQAKGN-LTNY 117  
Qy 121 INAIINSKVSDAINRVSAIREVVSAINEKMLQQEQDQKAAVEQKQENQAAINTVAANQE 180  
Db 118 INTILNSKVSDAVNRVVAIREVVSAINEKMLAQEQADKALEAQIENQNAINTVAANKQ 177  
Qy 181 TIAQNTVALNTQQAQLEAQLNLAELTTAQDQKATLVAQKAAAEBAARQAAQAQAAEA 240  
Db 178 AIENKKAALATQRAQLEAQLLSAQLTTVQNERKASLTQAKAQAEEAARKAAEAQAQAAEA 237  
Qy 241 KARAERKALQQAQQAQAAANNNTQATDASDQQAADNTQAATQAGDSTE--QSAQAQVN 298  
Db 238 KAQEAQAQEAQSAQAQAA-----QVESATAPTETVOTQPTKPNLTATS 286  
Qy 299 NSDQESTTATEA-----QPSASSASTAA-----VAANTSSANTYPAGQ 336  
Db 287 SATTVATTATATNEPKVTQPSVVTKAVEAPKAVVSTTPRAVSKPVRSYDSSNTYPMGQ 346  
Qy 337 CTWGVKSLAPVWGYNGWGGQWAAASAAAGYRVGSTPSAGAVAW--NDGGYGHVAVVTG 394  
Db 347 CTWGAKSMAWGVWGYNGWGNQWASARAAGYSVGTTPRVGAVAVWYPYDGGYGHVAVVTS 406  
Qy 395 V-QGGQIQVQFANYAGNOSIGNYRGWFPN---GSVSVIYPN 431  
Db 407 VANNSSIQVMSNYAGNMSIGNYRGSPNPSAGSVYIYPN 447

RESULT 13  
ADU69524  
ID ADU69524 standard; protein; 447 AA.  
XX AC  
XX ADU69524;  
DT 10-FEB-2005 (first entry)  
DE S agalactiae hyperimmune serum reactive antigen seqid 219.  
XX  
XX immune stimulation; antigen; bacterial surface display;  
KW hyperimmune serum reactive antigen; vaccine; bacterial infection;  
KW antibacterial; infection.  
XX  
OS Streptococcus agalactiae.  
XX

PN WO2004099242-A2.  
XX 18-NOV-2004.  
XX 06-MAY-2004; 2004WO-EP004856.  
XX 07-MAY-2003; 2003EP-00450112.  
PR 28-NOV-2003; 2003EP-00450266.  
XX (INTE-) INTERCELL AG.  
PI Meinke A, Nagy E, Hanner M, Horky M, Kallenda S, Prustomersky S;  
XX WPI; 2004-821662/81.  
DR N-PSDB; ADU69307.  
XX  
PT New nucleic acid molecule encoding a hyperimmune serum reactive antigen,  
PT useful for the manufacture of a vaccine against Streptococcus agalactiae  
PT infection.  
XX  
XX Claim 14; SEQ ID NO 219; 221pp; English.  
PS  
XX  
CC The invention describes an isolated nucleic acid molecule encoding a  
CC hyperimmune serum reactive antigen or its fragment. Also described are: a  
CC vector comprising the nucleic acid molecule; a host cell comprising the  
CC vector; a hyperimmune serum-reactive antigen comprising a sequence  
CC encoded by the nucleic acid molecule and consisting of e.g., 85, 299, 467  
CC or 812 amino acids; fragments of hyperimmune serum-reactive antigens  
CC consisting of peptides comprising e.g., 76, 134, 221 or 576 amino acids;  
CC a process for producing a Streptococcus agalactiae hyperimmune serum  
CC reactive antigen; a process for producing a cell that expresses a S.  
CC agalactiae hyperimmune serum reactive antigen; a pharmaceutical  
CC composition, especially a vaccine, comprising the hyperimmune serum-  
CC reactive antigen or nucleic acid molecule; an antibody that binds at  
CC least to a selective part of the hyperimmune serum-reactive antigen; a  
CC hybridoma cell line, which produces the antibody; a method for producing  
CC the antibody; an antagonist that binds to the hyperimmune serum-reactive  
CC antigen; a method for identifying an antagonist capable of binding to the  
CC hyperimmune serum-reactive antigen; a method for identifying an  
CC antagonist capable of reducing or inhibiting the interaction activity of  
CC a hyperimmune serum-reactive antigen to its interaction partner; a  
CC process for in vitro diagnosing a disease related to expression of the  
CC hyperimmune serum-reactive antigen; and a process for in vitro diagnosis  
CC of a bacterial infection, especially a S. agalactiae infection. The  
CC hyperimmune serum reactive antigen is useful for isolating, purifying  
CC and/or identifying an interaction partner of the hyperimmune serum  
CC reactive antigen. The hyperimmune serum reactive antigen is useful for  
CC generating a peptide binding to the hyperimmune serum reactive antigen,  
CC where the peptide comprises anticalines, or for the manufacture of a  
CC functional nucleic acid comprising aptamers or spiegelmers. The nucleic  
CC acid molecule is useful for the manufacture of a functional ribonucleic  
CC acid comprising ribozymes, antisense nucleic acids or siRNA. The nucleic  
CC acid molecule, hyperimmune serum-reactive antigen or antibody is useful  
CC for the manufacture of a vaccine against S. agalactiae infection. This is  
CC the amino acid sequence of a Streptococcus agalactiae hyperimmune serum  
CC reactive antigen.  
XX  
SQ Sequence 447 AA;  
Query Match 58.7%; Score 1235; DB 8; Length 447;  
Best Local Similarity 60.1%; Pred. No. 2.4e-75;  
Matches 277; Conservative 43; Mismatches 97; Indels 44; Gaps 9;  
Qy 1 MKKRILSAVLVSGVTLSSATTLSSAVKADDFPAQIASQDSKINNLTAAQQAAQAQVNTIQ 60  
Db 1 MKKRILSAVLVSGVTLGTAAV--TVNADDFDSKIAATDSVINTLSGGQAAQNVTAIKG 58  
Qy 61 QVSLQTOQAQLEAENORLEAQSATLGQOIQTLSSKIVARNESLKQQAARSQAKSNAATSY 120  
Db 59 QVGALESQQSELEAQAQLEAVSQOLGOEIQTLNKNIVARNESLKQVRSQAKGN-LTNY 117  
Qy 121 INAIINSKVSDAINRVSAIREVVSAINEKMLQQEQDQKAAVEQKQENQAAINTVAANQE 180  
XX

Db 118 INTILNSKSVSDAVNRVVVAIREVVVSANEKMLAQEADKALEAKQIENQNAINTVAANKQ 177  
Qy 181 TIAQNTNALNTQOALEAQLNLOAELTTAODOKATLVAKQAAAEAAARQAAAAQAAAEA 240  
Db 178 AIENKKAALATQRAQLEAQLLSAQLTTVQNEKASLIQAKAQAEAAARQAAEA 237  
Qy 241 KAAAEAKALOEQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGDSTE--QSAAQAVN 298  
Db 238 KAQAQAKAQAESVAKAQAAA-----QVESATAPTETVOTQPRTEIKPSNLTATS 286  
Qy 299 NSDQESTTATEA-----QPSASSASTAA-----VAANTSSANTYPAGQ 336  
Db 287 SATTVATTATATNEPKVTQPSVVTKAVEAPKAVVSTPRAVSKPVVRSYDSSNTYPMGQ 346  
Qy 337 CTWGVKSLAPWGVNGYWGNGQWAAASAAAAGYRVGSTPSAGAVW--NDGGYGHVAVTG 394  
Db 347 CTWGAKSMAWGVNGYWGNGWQNGASARAAGYSVGTTPRVGAVAWPYDGGYGHVAVVTS 406  
Qy 395 V-QGGQIQOQEANYAGNOSIGNYRGWFNP---GSVSYIYPN 431  
Db 407 VANNSSIQVNESYAGNMSIGNYRGSFNPSAGSGSVYIYPN 447  
  
RESULT 14  
ADV88392  
ID ADV88392 standard; protein; 447 AA.  
AC ADV88392;  
XX  
DT 24-FEB-2005 (first entry)  
XX  
DE Streptococcus agalactiae protein sequence, SEQ ID 786.  
XX  
KW Antibacterial; Vaccine; bacterial infection.  
XX  
OS Streptococcus agalactiae.  
XX  
PN FR2824074-A1.  
XX  
PD 31-OCT-2002.  
XX  
PF 26-APR-2001; 2001FR-00005642.  
XX  
PR 26-APR-2001; 2001FR-00005642.  
XX  
PA (INSP ) INST PASTEUR.  
XX  
PA (CNRS ) CNRS CENT NAT RECH SCI.  
XX  
PI Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
PI Zouina M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;  
XX WPI; 2004-101891/11.  
XX  
XX Genomic nucleotide sequences encoding polypeptides of Streptococcus  
XX agalactiae for the development of vaccines, diagnostic tools, DNA chips  
XX and identification of therapeutic targets.  
XX  
XX Claim 6; SEQ ID NO 786; 2687pp; French.  
XX  
XX The present invention relates to novel Streptococcus agalactiae  
XX nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;  
XX ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.  
XX agalactiae involved in the synthesis of amino acids, cell membranes, and  
XX intermediate (central) metabolism, energetic metabolism, fatty acid and  
XX phospholipid metabolism, nucleotide metabolism including purines,  
XX pyrimidines and/or nucleosides, regulatory functions, replication,  
XX transcription, translation, protein transport, adaptation to atypical  
XX conditions, sensitivity to medicines and/or analogues, functions related  
XX to transposons, biosynthesis of cofactors, prosthetic groups and  
XX transporters, cell membrane proteins and cellular machinery. (I) are  
XX useful for the detection and/or amplification of nucleic acids.  
XX Pharmaceutical composition comprising (I) or (II) are useful for  
XX treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is

CC equivalent for the present basic patent FR2824074A1. WO200292818A2  
CC contains 6617 sequence whereas the present patent only contains 2344  
CC sequences.  
XX  
SQ Sequence 447 AA;  
  
Query Match 58.7%; Score 1235; DB 8; Length 447;  
Best Local Similarity 60.1%; Pred. No. 2.4e-75;  
Matches 277; Conservative 43; Mismatches 97; Indels 44; Gaps 9;  
  
Qy 1 MKKRILLSAVLVSGVTLSATTLSAVKADDFDAQIAGDSKINNLTAAQQAAQAQVNTIQG 60  
Db 1 MKKRILLSAVLVSGVTLLGTAAV--TVNADDFDSKIATDTSVINTLSGQAAAAQNVTAIKG 58  
  
Qy 61 QVSALOTOQAELQAEVNRLEAQSATILCOQIOTLSSKIVARNESLKKQARSQAOKSNAATSY 120  
Db 59 QVGALSESQSELEAQAQNAQLSAVSSQQLGQEIQTLSNKIVARNESLKKQVRSQAQGN-LTNY 117  
  
Qy 121 INAIINSKVSDAINRVSAREVVVSANEKMLQOQEOBKAAVEKQOQENQAAINTVAAAOE 180  
Db 118 INTILNSKSVSDAVNRVVVAIREVVVSANEKMLAQEADKALEAKQIENQNAINTVAANKQ 177  
  
Qy 181 TIAQNTNALNTQOALEAQLNLOAELTTAODOKATLVAKQAAAEAAARQAAAAQAAAEA 240  
Db 178 AIENKKAALATQRAQLEAQLLSAQLTTVQNEKASLIQAKAQAEAAARQAAEA 237  
  
Qy 241 KAAAEAKALOEQAQAQAANNNNTQATDASDQAAAAADNTQAAQTGDSTE--QSAAQAVN 298  
Db 238 KAQAQAKAQAESVAKAQAAA-----QVESATAPTETVOTQPRTEIKPSNLTATS 286  
  
Qy 299 NSDQESTTATEA-----QPSASSASTAA-----VAANTSSANTYPAGQ 336  
Db 287 SATTVATTATATNEPKVTQPSVVTKAVEAPKAVVSTPRAVSKPVVRSYDSSNTYPMGQ 346  
  
Qy 337 CTWGVKSLAPWGVNGYWGNGQWAAASAAAAGYRVGSTPSAGAVW--NDGGYGHVAVTG 394  
Db 347 CTWGAKSMAWGVNGYWGNGWQNGASARAAGYSVGTTPRVGAVAWPYDGGYGHVAVVTS 406  
  
Qy 395 V-QGGQIQOQEANYAGNOSIGNYRGWFNP---GSVSYIYPN 431  
Db 407 VANNSSIQVNESYAGNMSIGNYRGSFNPSAGSGSVYIYPN 447  
  
RESULT 15  
ADV81808  
ID ADV81808 standard; protein; 447 AA.  
XX  
AC ADV81808;  
XX  
DT 24-FEB-2005 (first entry)  
XX  
DE Streptococcus agalactiae protein, SEQ ID 2949.  
XX  
KW Antibacterial; vaccine; bacterial infection.  
XX  
OS Streptococcus agalactiae.  
XX  
PN WO200292818-A2.  
XX  
PD 21-NOV-2002.  
XX  
PF 26-APR-2002; 2002WO-IB003059.  
XX  
PR 26-APR-2001; 2001FR-00005642.  
XX  
XX (INSP ) INST PASTEUR.  
XX (CNRS ) CNRS CENT NAT RECH SCI.  
XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;  
XX Zouina M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;  
XX WPI; 2004-101891/11.  
XX





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:20:08 ; Search time 50.8337 Seconds  
(without alignments)  
815.787 Million cell updates/sec

Title: US-10-797-821-30  
Perfect score: 2104  
Sequence: 1 MKKRILSAVLVSGVTLSSAT.....SIGNRWGFNGSVSYIYPN 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1057.5	50.3	392	2 G95258	secreted 45 kd pro
2	1057.5	50.3	392	2 B98124	general stress pro
3	676	32.1	461	2 JN0097	secreted 45K prote
4	661.5	31.4	456	2 E86903	hypothetical prote
5	547	26.0	507	2 S05542	hypothetical prote
6	359.5	17.1	398	2 AC1763	peptidoglycan lyti
7	346	16.4	401	2 A11387	cell wall-binding
8	317.5	15.1	473	2 F70031	conserved hypotet
9	268	12.7	581	2 E75383	cell wall binding
10	259	12.3	436	2 AH1387	cell wall binding
11	257.5	12.2	437	2 AB1763	cell wall-binding
12	256.5	12.2	461	2 H84099	conserved hypotet
13	238.5	11.3	528	2 B75310	tolA protein - Esc
14	234.5	11.1	421	2 JV0057	membrane spanning
15	230	10.9	394	2 F90725	membrane spanning
16	230	10.9	394	2 G85576	probable tail fibe
17	227.5	10.8	971	2 B90835	probable tail fibe
18	227.5	10.8	973	2 C85693	hypothetical prote
19	225.5	10.7	166	2 C90029	M5 protein precurs
20	223.5	10.6	492	2 A28616	cell wall surface
21	223.5	10.6	476	2 E95206	hypothetical prote
22	222.5	10.6	265	2 B89837	hypothetical prote
23	222.5	10.6	267	2 F90028	hypothetical tail fibe
24	221.5	10.5	1122	2 G44887	probable tail fibe
25	221	10.5	392	2 F96937	cell wall-associat
26	218	10.4	255	2 G90061	hypothetical prote
27	218	10.4	1156	2 T34852	probable secreted
28	214.5	10.2	483	2 A26297	M6 protein - Strep
29	213	10.1	688	2 AB3179	conserved hypotet

30	212	10.1	536	2 D84325	Htr17 transducer (
31	211	10.0	1528	2 A60338	surface antigen A
32	210.5	10.0	539	2 A28549	M24 protein precu
33	210.5	10.0	1173	2 T30608	proteophosphoglyc
34	209.5	10.0	1773	2 A81937	IGA-specific metal
35	208.5	9.9	1566	2 A43607	cell surface anti
36	208	9.9	643	2 H84305	Htr8 transducer [i
37	207.5	9.9	1128	2 T30296	R27-2 protein - Tr
38	207.5	9.9	2481	2 D90011	FmtB protein [impo
39	206.5	9.8	1589	2 T13606	hypothetical prote
40	206	9.8	1081	2 S66736	transcription acti
41	205.5	9.8	564	2 A60115	M protein precurs
42	205	9.7	1815	2 C81169	IGA-specific metal
43	204.5	9.7	1561	2 S61314	IGA-specific metal
44	204	9.7	436	2 S30284	M protein precurs
45	203	9.6	351	2 B97273	uncharacterized pr

ALIGNMENTS

RESULT 1

G95258  
secreted 45 kd protein [imported] - Streptococcus pneumoniae (strain TIGR4)  
C:Species: Streptococcus pneumoniae  
C:Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C:Accession: G95258  
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umavam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A:Reference number: A95000; MUID:21357209; PMID:11463916  
A:Accession: G95258  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-392 <R>  
A:Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AE005672;  
A:Experimental source: strain TIGR4  
C:Genetics:  
A:Gene: SP2216

Query Match	50.3%	Score 1057.5;	DB 2;	Length 392;
Best Local Similarity	50.3%;	Pred. No. 6.3e-44;		
Matches	219;	Conservative	73;	Mismatches 92;
			Indels	51;
			Gaps	5;
Qy	1	MKKRILSAVLVSGVTLSSATLSAVKADDFDAQIASODSKINNLTACQQAQAQVNTIQG	60	
Db	1	MKKRILASLLSTVMVSVQAVLTAAHAETDDDKIAAQDNKISNLTAQQQEAQKQVDQIQE	60	
Qy	61	QVSALQTOQAELQAEORLEAQSATLGGQITLSSKIVARNESLKQAARSAQSKNAATSY	120	
Db	61	QVSAIQAEQNLQAEQNDRLQAESKKEGTELSKNIVSRNQSLQKARSQATNGAVTSY	120	
Qy	121	INAIINSKVSVDIAINRVSAIREVVVANEKMLQQEQDKAAVEQKQENQAANTVAANQE	180	
Db	121	INTVNSKSITEAIRVAAEMSEIVSANNKMLEQQKADKKAISEKQVANDNAINVIANQQ	180	
Qy	181	TIAQNTVALNTQQAQLEAAQNLQAEALTAAQDQKATLVAQKAAAEAAEQAAAQAAAEA	240	
Db	181	KLADDAQALTYTKAELKAAELSLAAEKATAGEKASLLEQKAAAEAAEAAYK-	239	
Qy	241	KAAAEAKALQEAQAQAQAANNNTQATDASDQQAADNTQAAQGDSTEQSAQAQAVNNS	300	
Db	240	-----EKASQQAQSVLASNTNLTAQVAVSESAAA	270	
Qy	301	DQESTTTATEAQPSASSASTAAVAANTSSANTYPAGQCTGWGVKSLAPVGNVWNGGQAA	360	
Db	271	-----TNASSYPIGECTGWGVKTLAPWAGDYWGGAQWAT	315	
Qy	361	SAAAGRVGTSAGAVAVVNDGGYGHVAVVTQVG-QGIVQVEANYAGNQSGNVRGW	419	

Db 316 SAAAAGFRTGTPQVGAIAACWNDGGYGHVAVVTAVESTTRIQVESNYAGNRTIGNHRGW 375  
QY 420 FNP-----GSVSYIY 429  
Db 376 FNPTTSEGFTYIY 390  
RESULT 2  
B98124  
General stress protein GSP-781 [imported] - Streptococcus pneumoniae (strain R6)  
C;Species: Streptococcus pneumoniae  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C;Accession: B98124  
R;Hoekings, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; DeHoff, B.S.; E  
s, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A;Reference number: A97872; MUID:21429245; PMID:11544234  
A;Accession: B98124  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-392 <KUR>  
A;Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AE007317;  
C;Genetics:  
A;Gene: gsp-781

Query Match 50.3%; Score 1057.5; DB 2; Length 392;  
Best Local Similarity 50.3%; Pred. No. 6.3e-44;  
Matches 219; Conservative 73; Mismatches 92; Indels 51; Gaps 5;  
QY 1 MKKRIISAVLVSGVTLSSATTLTSAVKADDFDAQIASODSKINNLTAAQQAQAQVNTIQG 60  
Db 1 MKKKILASLLSTVWVSQVAVLTAAHAEETDDKIAAQDNKISNLTAAQQAQAQVNTIQE 60  
QY 61 QVSALOTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQQAQAQAQAQAQA 120  
Db 61 QVSAIQAEQNLQAEORLEAQSATLGGQIOTLSSKIVARNESLKQQAQAQAQAQAQA 120  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQEDQKAQVQEQQAQAQAQAQAQA 180  
Db 121 INTVNSKSIETATSRVAAMSEIVSANNKMLEQOQEDQKAQVQEQQAQAQAQAQAQA 180  
QY 181 TIAQNTNALNTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQQAQAQAQA 240  
Db 181 KLADDAQALTTKQAELEAELSLAAEKATAEAGEKASLLEQKAAAEARAAAEAAAYK- 239  
QY 241 KAAAEAKALQEAQA 300  
Db 240 -----EKRAQQQSVLASANTNLTAQVQAVSESAAA----- 270  
QY 301 DQESTTTATEAOPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVWVNGWGGOWAA 360  
Db 271 ----PVRAKVRPTVS-----TNASSYPIGECTGWGKTLAPWAGDWGNGAQWAT 315  
QY 361 SAAAAGRVSTPSAGAVAVWNDGGYGHVAVVTQVQ-GQIQVQEAQYAGNQSIGNVRGW 419  
Db 316 SAAAAGFRTGTPQVGAIAACWNDGGYGHVAVVTAVESTTRIQVESNYAGNRTIGNHRGW 375  
QY 420 FNP-----GSVSYIY 429  
Db 376 FNPTTSEGFTYIY 390

RESULT 3  
JN0097  
secreted 45K protein precursor - Lactococcus lactis  
C;Species: Lactococcus lactis  
C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999  
C;Accession: JN0097  
R;van Asseldonk, M.; Rutten, G.; Otman, M.; Siezen, R.J.; de Vos, W.M.; Simons, G.  
Gene 95, 155-160, 1990

A;Title: Cloning of usp45, a gene encoding a secreted protein from Lactococcus lactis ssp  
A;Reference number: JN0097; MUID:91071599; PMID:2123812  
A;Accession: JN0097  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-461 <VAN>  
A;Cross-references: UNIPARC:UPI000016D72B; GB:M35374  
Query Match 32.1%; Score 676; DB 2; Length 461;  
Best Local Similarity 35.2%; Pred. No. 1.2e-25;  
Matches 169; Conservative 104; Mismatches 139; Indels 68; Gaps 13;  
QY 1 MKKRIISAVLVSGVTLSSATTLTSAVKADDFDAQIASODSKINNLTAAQQAQAQAQVNTIQG 60  
Db 1 MKKKIISAILMSVILSAAPLSGVADT-NSDIAKODATISSAQSAKAQAQAQVDSLQS 59  
QY 61 QVSALOTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQQAQAQAQAQAQA 120  
Db 60 KVDLSQQKQTSKTAQIAKIESEAKALNAQIATLINESIKERTKTLEAQAQAQAQAQAQA 119  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOQEDQKAQVQEQQAQAQAQAQAQA 180  
Db 120 MDAVWNSKSLTDVIQVTAITATVSSANKQMLEQOQEDQKAQVQEQQAQAQAQAQA 179  
QY 181 TIAQNTNALNTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQQAQAQAQAQA 240  
Db 180 SLDQSAQELTSQAELKVATLNTQIATIAQDQKQALLDEKAAAEAKQAQAQAQAQA 239  
QY 241 KAAAEAKALQEAQA 295  
Db 240 Q-----QKEAAQA 291  
QY 296 AVNNSDQESTTATEAOPSASSASTAAVAANTSS-----ANTYP 333  
Db 292 SSNSSS 351  
QY 334 AGQCT---WG-----VKSLAPVWVNGWGGOWAAASAAAAG--YRVGSTPSAGAVAV 380  
Db 352 GGCTDYVWQYFAAQGIYIRNIMP-----GNGQWASNGPAQGVHLHVGAAP--GVIAS 403  
QY 381 -----WNDGGYGHVAVVTQVQ-GQIQVQEAQYAGNQSIGNVRGWFPNGSVSYIYPN 431  
Db 404 SPSADPFGVANSYPGHVAIVKSVNSDGTITIKEGY-GTTWNGHER-TVSASGVTFMLPN 461  
RESULT 4  
E86903  
hypothetical protein usp45 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)  
C;Species: Lactococcus lactis subsp. lactis  
C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 09-Jul-2004  
C;Accession: E86903  
R;Bolotin, A.; Wincker, P.; Mager, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich  
Genome Res. 11, 731-753, 2001  
A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp  
A;Reference number: A86625; MUID:21235186; PMID:11334741  
A;Accession: E86903  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-456 <STO>  
A;Cross-references: UNIPROT:Q9CDJ1; UNIPARC:UPI000000C6BFC; GB:AE005176; PID:g12725296; P  
A;Experimental source: strain IL1403  
C;Genetics:  
A;Gene: usp45  
Query Match 31.4%; Score 661.5; DB 2; Length 456;  
Best Local Similarity 34.6%; Pred. No. 6e-25;  
Matches 165; Conservative 108; Mismatches 137; Indels 67; Gaps 14;  
QY 1 MKKRIISAVLVSGVTLSSATTLTSAVKADDFDAQIASODSKINNLTAAQQAQAQAQVNTIQG 60  
Db 1 MKKKIISAILMSVILSAAPLSGVADT-NSDIAKODATISSAQSAKAQAQAQVDSLQS 59  
QY 61 QVSALOTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQQAQAQAQAQAQA 120

```

Db      358  STNNNNNNNTVTPAPPTPTTPAPAPAPNGSUNGALIAEAYKVIQTPYV---WGGKD 414
Qy      356  -----GOWASASAAGYRVG-STPSAGAVAVMND--GGYGHVAY 391
Db      415  PSGFDCSGFTRYVYLQVTGRDIGMTVPQESAGTKISVQAKAGDLLFWGSAGGTYHVAI 474
Qy      392  VTGVOGQGIQOEANYAGNOSIGNYRGWFNP 422
Db      475  SLG--GGQ-YIHAPQPGENVKVGSVQ--WYTP 501

RESULT 6
AC1763
pepidoglycan lytic protein P45 [imported] - Listeria innocua (strain Clip126)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AC1763
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.
; Dominguez-bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitourn
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1763
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <GLA>
A:Cross-references: UNIPROT:Q927Y8; UNIPARC:UPI00000CC935; GB:AL592022; PIDN:C
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: spl

Query Match 17.1%; Score 359.5; DB 2; Length 398;
Best Local Similarity 27.2%; Pred. No. 1.4e-10;
Matches 114; Conservative 88; Mismatches 146; Indels 71; Gaps 16

Qy      1  MKKEILSAV-LVSGVTLLSATTLNAVKADDFDAQIASODSKNNLTAQOAAQAQVNTIQ 59
Db      1  MKQNTFAISLAAVISUTPAFTNVFA--DVNTDIQNQDKKINDIKSKRTGQSDLSLV 58

```

Qy	59	ADLEKAQEAKSLQGEFDQTKELKQLNQDIKDINERIKETVILKERARAWQKTSNSNA	118
Db		ADLEKAQEAKSLQGEFDQTKELKQLNQDIKDINERIKETVILKERARAWQKTSNSNA	118
Qy	120	YINAIINSKSVSDAINRVSAITREVVSAANEKMLQQEQED-----KAAVEQKQEQENQAAI	172
Db		YINAIINSKSVSDAINRVSAITREVVSAANEKMLQQEQED-----KAAVEQKQEQENQAAI	172
Qy	119	YLEVILDAENLSDLVGRVSANVLVDSKSILEDQDQKDEKALKTKQTAVKKKQEEQATAI	178
Db		YLEVILDAENLSDLVGRVSANVLVDSKSILEDQDQKDEKALKTKQTAVKKKQEEQATAI	178
Qy	173	NTVAANQETIAQNTNALNMQQAQLEAAQLNQAEILTTAQDQKATILVAQKAAAEAAARQAA	232
Db		NTVAANQETIAQNTNALNMQQAQLEAAQLNQAEILTTAQDQKATILVAQKAAAEAAARQAA	232
Qy	179	HEFEAQO-----NKTEAQAKAEKAIIVAQLAADQSAENEKAGLVSE--DKAAKE--	226
Db		HEFEAQO-----NKTEAQAKAEKAIIVAQLAADQSAENEKAGLVSE--DKAAKE--	226
Qy	233	AAQAAAAEAKAAAEAKALQEQAAQAAQAAANNNNTQATDASDQQAADNTQAAQTGDPSTES	292
Db		AAQAAAAEAKAAAEAKALQEQAAQAAQAAANNNNTQATDASDQQAADNTQAAQTGDPSTES	292
Qy	227	-----ATARATALRE-----AIDANVGQQT--NTNA--SSNSKTSN	259
Db		-----ATARATALRE-----AIDANVGQQT--NTNA--SSNSKTSN	259
Qy	293	AAQAVNNSDQESTTATEAQPSASASTAAVAANTSSANTYPAG-----QCTMGVKSIL--	344
Db		AAQAVNNSDQESTTATEAQPSASASTAAVAANTSSANTYPAG-----QCTMGVKSIL--	344
Qy	260	KVESTNNSAEPSA---ATPSSGGVYSAMIAAARALQPKYSLGATGPSAFDCS-GFTSYA	314
Db		KVESTNNSAEPSA---ATPSSGGVYSAMIAAARALQPKYSLGATGPSAFDCS-GFTSYA	314
Qy	345	--APWGVNYNGNGGOWAASAAAAAGYRVGSTPSAGAVAVWN-DGGYGHVAVYTVGVGGQI	400
Db		--APWGVNYNGNGGOWAASAAAAAGYRVGSTPSAGAVAVWN-DGGYGHVAVYTVGVGGQI	400
Qy	315	FRAAGVSLPRTSGGQYAAAASKIS---ASQAKPGDLVFFNYGGGIAHVGIIYVG--GGQM	367
Db		FRAAGVSLPRTSGGQYAAAASKIS---ASQAKPGDLVFFNYGGGIAHVGIIYVG--GGQM	367
RESULT 7			
AII387			
peptidoglycan lytic protein P45 [imported] - Listeria monocytogenes (strain EGGS)			
C/Species: Listeria monocytogenes			
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004			

RESULT 7  
A11387  
peptidoglycan lytic protein P45 [imported] - *Listeria monocytogenes* (strain EGD-e)  
C:Species: *Listeria monocytogenes*  
C:Date: 27-Nov-2001 #sequence revision 27-Nov-2001 #text change 09-Jul-2004



```
Db 40 LAGTDTLSRLQQLRQLIEQRILSVQOKQLTEVRRLRLQLNAQQAALDRDLALTSE 99
Qy 62 VSALQTQOAELOAENRLEAQSATLQ---QIQTLSKIVARNESLQQAARSQKS---N 115
Db 100 VTDLENELADVLA---RVTAATRALRETEAQIRVTRSQV-----EALKVDABAVMKALYRA 152
Qy 116 AATSYINAINSKSVSDAINRVSAIREVVSANEMKLOQ-----QSDKAAY 161
Db 153 RNTQYMLLSQSNISDMLIRLDYANMAGQNVNVEMLRGAAAEALTQQLRQRQSDAL 212
Qy 162 EQKQENQ-----AAINTVAANQETIAQNTNALNTQQAQ----- 195
Db 213 RGLQEQOQTKLAELDRDRTRQADMLAELQRSAQCGQAVVTRTQQAALTAQTIQSLVGNV 272
Qy 196 -----LEAQLNLOAE-----LTTAQQOKATLVA-- 219
Db 273 VRETRLEERRRLEERRRREARRIRAEQERKAREAAEA 332
Qy 220 --QVAAAEARQAAAQAAAEAKAAEAKALQE-----QAAQAAANNNTQ----- 265
Db 333 ARQKALAE--ARAAQAKVAAEAKARAEAKARAEAAQAAQAAQAAQAAQAAQAAE 390
Qy 266 --ATDASDQAAAADNTQAAQTGSTQSAQAQVN-----NSDQESTTATE 309
Db 391 AARAQAQQAQAAAARAEAAQARAQAQQAQQAQKVRQAVAREQRDLQTOQOQTAQOE 450
Qy 310 AQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGGQAAASAAAAGYRV 369
Db 451 KQLAAELAPLPAISRDGL--FPLP-----GGKVAAP-----YGTSGAQWEVLSGASGRA 499
Qy 370 GSTPSAGAVA-----VWMDGGYGHVAYVTGVQGGQIQV 402
Db 500 VAAQTGNVIAAYAAALGWILLDHGNSVITGYGLQDTLVEV 542

RESULT 10
AH1387
cell wall binding proteins homolog lmo2504 [imported] - Listeria monocytogenes (strain E
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AH1387
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Krest, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Varquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AH1387
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-436 <GLA>
A;Cross-references: UNIPROT:Q8Y4E2; UNIPARC:UPI0000055221; GB:NC_003210; PIDN:CAD00582.1
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo2504

Query Match 12.3%; Score 259; DB 2; Length 436;
Best Local Similarity 23.3%; Pred. No. 9.4e-06;
Matches 98; Conservative 86; Mismatches 160; Indels 76; Gaps 14;

Qy 6 LSAVLVSG-VTLSSATLTSV-----KADDFDAQIASQDSKINNLTAAQQAQAV 55
Db 11 LSLIIISAPLGAADINDMKQKNEIEQKSIDKNNLSKSELNHLSEAEKDAKEL 70
Qy 56 NTIQGV-----SALQTQOAELOAENRLEAQSATLQGIQITLSSKIVARNESLQQAARS 111
Db 71 ESLMKSLLDTNKKLKEQEDKVSSSEKLGK----KLQKEMEKLRNDRQKVLNRAAI 126
Qy 112 QKSNAASTSYINAINSKSVSDAINRVSAIREVVSANEMKLOQ---EQDKAAVEQKQENQ 169
```

```
Db 127 QTTGTATSYLDMIFEADDFKELVDVTRTVVSAIVKADQNIMQDKDDKLVKVAESTSEKK 186
Qy 170 AA-INTVAANQETIAQNTNALNTQOAELOAENRLEAQSATLQGIQITLSSKIVARNESLQQAARS 228
Db 187 LENLKVLALELVSKNNMESQKQKNDLVWALAN-KKDLT--KSEQTLLASEQGALTDEE 243
Qy 229 RQAAAQAAAQAAAEAKALQEOAAQAAANNTQATDASDQQAADNTQAAQTGDS 288
Db 244 KRLASNTAGEKAKQEAIAKAAEKKM-----QEAASAKSA----- 281
Qy 289 TEQSAAQAVNNSDQESTTATEAQPSASSASTAAVAANTSSANTYPA-GCCTWGVKSLA-P 346
Db 282 -----AVVKQPSSSNEATEITVSSGGQFIKPASGILTSGFSEINP 323
Qy 347 WVGNYWGN-----GGQWAAASAAAAGYRVGSTPSAGAVAVWMDGGYGHVAYVTGVQGGQ 399
Db 324 VTGKYSHKGQDIAGGGTIVTVAASGTVVPSFGSGSGSGF---GGYGVVVKIDHNGNGFQ 380

RESULT 11
AB1763
cell wall binding protein homolog lin2647 [imported] - Listeria innocua (strain Clip1126
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AB1763
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Krest, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Varquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AB1763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-437 <GLA>
A;Cross-references: UNIPROT:Q927Y9; UNIPARC:UPI000000CC934; GB:ALU592022; PIDN:CAC97874.1;
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin2647

Query Match 12.2%; Score 257.5; DB 2; Length 437;
Best Local Similarity 24.2%; Pred. No. 1.1e-05;
Matches 106; Conservative 84; Mismatches 137; Indels 111; Gaps 18;

Qy 6 LSAVLVSGVTLSSATLTSVAKAD-----DFDAQIASQDSKINNLTAAQ 47
Db 11 LSLIIIS-----APLTSVKAESINDMKQKSEIEQKSELNKLNTKSELNHLSEAE 63
Qy 48 QQA--QAQVNTTIQGVSAALQTOQAELQAEQENRLEAQSATLQGIQITLSSKIVARNESL 104
Db 64 KDAAKELESLLNSIDETNKKLKEQEDKVDSSEKLGK---KLKEIEKLRNDRQKVL 119
Qy 105 KQARSQAKNAATSYINAINSKSVSDAINRVSAIREVVSANEMKLOQEQDQAAVE-- 162
Db 120 DSRARAIQTTGTATSYLDMIFEADDFKELIDRVVWSAIVKADQNIMQDKDDQKLVKA 179
Qy 163 ----QKQENQAAINTVAANQETIAQNTNALNTQOAELOAENRLEAQSATLQGIQITLSSKIVARNESL 218
Db 180 ENSEKLEN---LKVLALELVSKNNMESQKQKNDLVWALAN-KKDLT--KSEQTLLT 233
Qy 219 AQAAAEEAARQAAAQAAAEAKAAEAKALQEOAAQAAANNTQATDASDQQAAD 278
Db 234 NEQALSDQEKKLASNTAGEKAKQEAIAKAAEKKM-----QEAASAKN 276
Qy 279 NTOAAQTGDSSTEGSAAQAVNNSDQESTTATEAQPS--SSASTAAVAANTSSANTY--PA 334
Db 277 ATVAAQ-----QPSSVTSSAGTGTATDVTSSGGGQFIKPA 310
Qy 335 -GCCTWGVKSLA-PWVGNWGN-----GGQWAAASAAAAGYRV---GSTPSAGAVAV 381
Db 311 SGMLTSGFSERTNPVTGKYSHKGQDIAGGGTIVTVAASGTVVPSFGSGFATGS----- 363
```





Db 119 QAEEAAKQAEKQKQAEAAAKAAADAKAKAEADDKAAEEAAKKAADAKKKAEEAAKA 178

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:07:53 ; Search time 295.878 Seconds  
(without alignments)  
1027.730 Million cell updates/sec

Title: US-10-797-821-30  
Perfect score: 2104  
Sequence: 1 MKKRILSAVLNVSGVTLSAT.....SIGNYRGWFGNSVSIYPN 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2104	100.0	431	2	Q9AG98_STRMU
2	2100	99.8	431	2	Q938V3_STRMU
3	2100	99.8	431	2	Q8DWM3_STRMU
4	2096	99.6	431	2	Q938V0_STRMU
5	2083.5	99.0	432	2	Q938V2_STRMU
6	2074.5	98.6	432	2	Q938V1_STRMU
7	1235	58.7	447	2	Q9AKA4_STRAG
8	1235	58.7	447	2	Q8E2H1_STRAS
9	1235	58.7	447	2	Q8E7X9_STRAS
10	1094.5	52.0	485	2	Q5M212_STRT1
11	1093	51.9	474	2	Q5M6K4_STRT2
12	1092.5	51.9	398	2	Q5XEL1_STR66
13	1092.5	51.9	398	2	Q7CQK7_STRP8
14	1092.5	51.9	398	2	Q8P318_STRP3
15	1085.5	51.6	398	2	Q9A128_STRPY
16	1057.5	50.3	392	2	Q8DMY4_STRR6
17	1057.5	50.3	392	2	Q97N55_STRPN
18	947	45.0	211	2	Q9ZAS7_STRMU
19	669	31.8	461	1	USP45_LACLC
20	661.5	31.4	456	2	Q9CDJ1_LACLA
21	594	28.2	524	2	Q9K2J9_ENTFC
22	579	27.5	516	1	P54_ENTFC
23	563	26.8	576	2	Q9KJ33_ENTHR
24	554.5	26.4	482	2	Q5MSM6_STRT2
25	549.5	26.1	470	2	Q8RQE3_9LACT
26	531.5	25.3	482	2	Q931K4_ENTFA
27	518	24.6	461	2	Q56SA7_STRTR
28	507	24.1	449	2	Q931K3_ENTFA
29	435	20.7	211	2	Q8DVU8_STRMU
30	417.5	19.8	544	2	Q840X3_STRMU
31	416.5	19.8	226	2	Q93RG6_STRIT

32	415.5	19.7	544	2	Q840V8_STRMU	Q840V8 streptococc
33	412.5	19.6	544	2	Q840W6_STRMU	Q840W6 streptococc
34	410.5	19.5	544	2	Q8DUR7_STRMU	Q8dur7 streptococc
35	404.5	19.2	129	2	Q5M137_STRT1	Q5m137 streptococc
36	403	19.2	169	2	Q8E3F4_STRT3	Q8e3f4 streptococc
37	395	18.8	169	2	Q8DXT4_STRAS	Q8dxt4 streptococc
38	359.5	17.1	398	2	Q927Y8_STRISN	Q927y8 listeria in
39	346	16.4	401	2	Q71WS3_LISMP	Q71ws3 listeria mo
40	346	16.4	401	2	Q9RE04_LISMO	Q9re04 listeria mo
41	337.5	16.0	604	2	Q6M552_CORGL	Q6m552 corynebacte
42	335	15.9	430	2	Q81HV2_BACCR	Q81hv2 bacillus ce
43	331.5	15.8	600	2	Q8NOA0_CORGL	Q8nga0 corynebacte
44	322	15.3	453	2	Q8G539_BIFLO	Q8g539 bifidobacte
45	321.5	15.3	427	2	Q73DG0_BACCI	Q73dgo bacillus ce

ALIGNMENTS

RESULT 1  
Q9AG98\_STRMU PRELIMINARY; PRT; 431 AA.  
AC Q9AG98;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)  
DE Immunodominant glycoprotein IDG-60 (Glucan-binding protein B).  
GN Name=sagA;  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=GS-5;  
RX MEDLINE=21153617; PubMed=11254612;  
RX DOI=10.1128/IAI.69.4.2493-2501.2001;  
RA Chia J.S., Lee Y.Y., Huang P.T., Chen J.Y.;  
RT "Identification of stress-responsive genes in Streptococcus mutans by differential display reverse transcription-PCR.";  
RL Infect. Immun. 69:2493-2501(2001).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=GS-5;  
RX MEDLINE=21481977; PubMed=11598074;  
RX DOI=10.1128/IAI.69.11.6987-6998.2001;  
RA Chia J.S., Chang L.Y., Shun C.T., Chang Y.Y., Chen J.Y.;  
RT "A 60-kilodalton immunodominant glycoprotein is essential for cell wall integrity and the maintenance of cell shape in Streptococcus mutans.";  
RL Infect. Immun. 69:6987-6998(2001).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=GS-5;  
RX MEDLINE=21481971; PubMed=11598068;  
RX DOI=10.1128/IAI.69.11.6931-6941.2001;  
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J., Duncan M.J.;  
RT "Cloning of the Streptococcus mutans gene encoding glucan binding protein B and analysis of genetic diversity and protein production in clinical isolates.";  
RL Infect. Immun. 69:6931-6941(2001).

```

DR EMBL; AF338445; AAK08104.1; -; Genomic_DNA.
DR EMBL; AY046411; AAK94501.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
SQ SEQUENCE 431 AA; 44592 MW; 3BBE21FC5E47232E CRC64;

Query Match 100.0%; Score 2104; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 8.6e-86;
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSTATLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSTATLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60

QY 61 QVSALQTQQAELQAEORLEAQSATLGQQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
DB 61 QVSALQTQQAELQAEORLEAQSATLGQQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAIAINTVAANOE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAIAINTVAANOE 180

QY 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQQDKATLVAQKAAAEAAEAAQAAAAA 240
DB 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQQDKATLVAQKAAAEAAEAAQAAAAA 240

QY 241 KAAAEAKALQEQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDTSTEQSAAQVNN 300
DB 241 KAAAEAKALQEQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDTSTEQSAAQVNN 300

QY 301 DQESTTATEAOPSASSASTAAVAANTSSANTYPAGQCTGWKSLAPWVGNWNGGOWAA 360
DB 301 DQESTTATEAOPSASSASTAAVAANTSSANTYPAGQCTGWKSLAPWVGNWNGGOWAA 360

QY 361 SAAAAGRVGSTPSAGAVAVWNDGGYGHVAVYTVGVGGQIQVOEANYAGNQSIGNRGWF 420
DB 361 SAAAAGRVGSTPSAGAVAVWNDGGYGHVAVYTVGVGGQIQVOEANYAGNQSIGNRGWF 420

QY 421 NPGSVSYIYPN 431
DB 421 NPGSVSYIYPN 431

RESULT 2
Q938V3_STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5J32;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5J32;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
protein B and analysis of genetic diversity and protein production in

```

---

```

RT clinical isolates.";
RL EMBL; AY046410; AAK94500.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
SQ SEQUENCE 431 AA; 44620 MW; 464FE3B563FB7E51 CRC64;

Query Match 99.8%; Score 2100; DB 2; Length 431;
Best Local Similarity 99.8%; Pred. No. 1.3e-85;
Matches 430; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSSTATLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSTATLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60

QY 61 QVSALQTQQAELQAEORLEAQSATLGQQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
DB 61 QVSALQTQQAELQAEORLEAQSATLGQQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAIAINTVAANOE 180
DB 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAIAINTVAANOE 180

QY 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQQDKATLVAQKAAAEAAEAAQAAAAA 240
DB 181 TIAQNTNALNTQQAQLAEALNLQAEELTTAQQDKATLVAQKAAAEAAEAAQAAAAA 240

QY 241 KAAAEAKALQEQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDTSTEQSAAQVNN 300
DB 241 KAAAEAKALQEQAQAQAQAANNNTQATDASDQAAAAADNTQAAQTGDTSTEQSAAQVNN 300

QY 301 DQESTTATEAOPSASSASTAAVAANTSSANTYPAGQCTGWKSLAPWVGNWNGGOWAA 360
DB 301 DQESTTATEAOPSASSASTAAVAANTSSANTYPAGQCTGWKSLAPWVGNWNGGOWAA 360

QY 361 SAAAAGRVGSTPSAGAVAVWNDGGYGHVAVYTVGVGGQIQVOEANYAGNQSIGNRGWF 420
DB 361 SAAAAGRVGSTPSAGAVAVWNDGGYGHVAVYTVGVGGQIQVOEANYAGNQSIGNRGWF 420

QY 421 NPGSVSYIYPN 431
DB 421 NPGSVSYIYPN 431

RESULT 3
Q8DWM3_STRMU PRELIMINARY; PRT; 431 AA.
AC Q8DWM3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2004 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative secreted antigen GbpB/sagA; putative peptidoglycan
DE hydrolase.
GN Name=gbpB; OrderedLocusNames=SMU.22;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR EMBL; AE014855; AAN57811.1; -; Genomic_DNA.

```

```

DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 431 AA; 44620 MW; 2D1CA685248CCD3E CRC64;

Query Match          99.8%; Score 2100; DB 2; Length 431;
Best Local Similarity 99.8%; Pred. No. 1.3e-85;
Matches 430; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSSTLSSATTLSSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSSTLSSATTLSSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Qy 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
Db 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
Qy 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANQE 180
Qy 181 TIAQNTNALNTQQAQLAEALNLQAEITTTAQDQKATLVAQKAAAEAEARQAAAAQAAAEA 240
Db 181 TIAQNTNALNTQQAQLAEALNLQAEITTTAQDQKATLVAQKAAAEAEARQAAAAQAAAEA 240
Qy 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGUSTEQSAAQAQVNS 300
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGUSTEQSAAQAQVNS 300
Qy 301 DQESTTATEAQPSSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWAA 360
Db 301 DQESTTATEAQPSSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWAA 360
Qy 361 SAAAAGRVGSTPSAGAVAVWMDGGYGHVAYVTGVQGGQIQVQAEYAGNQSIGNYRGWF 420
Db 361 SAAAAGRVGSTPSAGAVAVWMDGGYGHVAYVTGVQGGQIQVQAEYAGNQSIGNYRGWF 420
Qy 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431

RESULT 4
Q938V0_STRMU
ID Q938V0_STRMU PRELIMINARY; PRT; 431 AA.
AC Q938V0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SSM3;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SSM3;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.;
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";

RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AY046414; AAK94504.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 431 AA; 44650 MW; 05D38D8DB4609F CRC64;

Query Match          99.6%; Score 2096; DB 2; Length 431;
Best Local Similarity 99.5%; Pred. No. 2e-85;
Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSSTLSSATTLSSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSSTLSSATTLSSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Qy 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
Db 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
Qy 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAAINTVAANQE 180
Qy 181 TIAQNTNALNTQQAQLAEALNLQAEITTTAQDQKATLVAQKAAAEAEARQAAAAQAAAEA 240
Db 181 TIAQNTNALNTQQAQLAEALNLQAEITTTAQDQKATLVAQKAAAEAEARQAAAAQAAAEA 240
Qy 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGUSTEQSAAQAQVNS 300
Db 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGUSTEQSAAQAQVNS 300
Qy 301 DQESTTATEAQPSSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWAA 360
Db 301 DQESTTATEAQPSSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGQWAA 360
Qy 361 SAAAAGRVGSTPSAGAVAVWMDGGYGHVAYVTGVQGGQIQVQAEYAGNQSIGNYRGWF 420
Db 361 SAAAAGRVGSTPSAGAVAVWMDGGYGHVAYVTGVQGGQIQVQAEYAGNQSIGNYRGWF 420
Qy 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431

RESULT 5
Q938V2_STRMU
ID Q938V2_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.;
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
RT protein B and analysis of genetic diversity and protein production in
RT clinical isolates.";

```

```
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL: AY046412; AAK94502.1; -, Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44648 MW; E769B2504AE50E9 CRC64;

Query Match 99.0%; Score 2083.5; DB 2; Length 432;
Best Local Similarity 99.1%; Pred. No. 7e-85;
Matches 428; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Qy 1 MKKRLSGLVSGVTLSATLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
Db 1 MKKRLSGLVSGVTLSATLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60

Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
Db 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNAAATSY 120

Qy 121 INAIINSKVSDAINRVSAIREVVSANEXMLQQEQDKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEXMLQQEQDKAAVEQKQENQAAINTVAANQE 180

Qy 181 TIAQNTNALNTQQAQLQAEALNLQAEELTTAQQDKATLVAQKAAAEAAQAQAAAE 240
Db 181 TIAQNTNALNTQQAQLQAEALNLQAEELTTAQQDKATLVAQKAAAEAAQAQAAAE 240

Qy 241 KAAAEAKALQEAQAQAQAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299
Db 241 KAAAEAKALQEAQAQAQAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300

Qy 300 SDQESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWGVNGWGGQWA 359
Db 301 SDQESTTATAAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWGVNGWGGQWA 360

Qy 360 ASAAAAGYRGVSTPSAGAVVNDGGYGHVAVYVTGVGGQIQVEANYAGNQSIGNYRGW 419
Db 361 ASAAAAGYRGVSTPSAGAVVNDGGYGHVAVYVTGVGGQIQVEANYAGNQSIGNYRGW 420

Qy 420 FNPGSVSYIYPN 431
Db 421 FNPGSVSYIYPN 432

RESULT 6
Q938V1_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
[1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
[2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RX MEDLINE=21481971; PubMed=11598068;
RX DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding

RT protein B and analysis of genetic diversity and protein production in
clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL: AY046413; AAK94503.1; -, Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44652 MW; 3F88ECB9A1F3BE4F CRC64;

Query Match 98.6%; Score 2074.5; DB 2; Length 432;
Best Local Similarity 98.8%; Pred. No. 1.8e-84;
Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 1 MKKRLSGLVSGVTLSATLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
Db 1 MKKRLSGLVSGVTLSATLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60

Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
Db 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNAAATSY 120

Qy 121 INAIINSKVSDAINRVSAIREVVSANEXMLQQEQDKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEXMLHQEQDKAAVEQKHQENQAAINTVAANQE 180

Qy 181 TIAQNTNALNTQQAQLQAEALNLQAEELTTAQQDKATLVAQKAAAEAAQAQAAAE 240
Db 181 TIAQNTNALNTQQAQLQAEALNLQAEELTTAQQDKATLVAQKAAAEAAQAQAAAE 240

Qy 241 KAAAEAKALQEAQAQAQAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299
Db 241 KAAAEAKALQEAQAQAQAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300

Qy 300 SDQESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWGVNGWGGQWA 359
Db 301 SDQESTTATAAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWGVNGWGGQWA 360

Qy 360 ASAAAAGYRGVSTPSAGAVVNDGGYGHVAVYVTGVGGQIQVEANYAGNQSIGNYRGW 419
Db 361 ASAAAAGYRGVSTPSAGAVVNDGGYGHVAVYVTGVGGQIQVEANYAGNQSIGNYRGW 420

Qy 420 FNPGSVSYIYPN 431
Db 421 FNPGSVSYIYPN 432

RESULT 7
Q9AKA4_STRAG PRELIMINARY; PRT; 447 AA.
AC Q9AKA4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PcsB protein precursor.
GN Name=PcsB;
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
[1]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=6313;
RX MEDLINE=21101799; PubMed=11157929;
RX DOI=10.1128/JB.183.4.1175-1183.2001;
RA Reinscheid D.J., Gottschalk B., Schubert A., Eikmanns B.J.,
RA Chhatwal G.S.;
RT "Identification and molecular analysis of PcsB, a protein required for
cell wall separation of group B streptococcus.";
RL J. Bacteriol. 183:1175-1183(2001).
DR EMBL; AJ277292; CAC28144.1; -, Genomic_DNA.
DR InterPro; IPR007921; CHAP.
```

```

DR InterPro: IPR009148; Siba.
DR Pfam: PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PSS0911; CHAP; 1.
KW Signal.
FT CHAIN 1 25 Potential.
FT CHAIN 26 447 PcsB protein.
SQ SEQUENCE 447 AA; 46681 MW; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 58.7%; Score 1235; DB 2; Length 447;
Best Local Similarity 60.1%; Pred. No. 2.7e-47;
Matches 277; Conservative 43; Mismatches 97; Indels 44; Gaps 9;

Qy 1 MKKRILSAVLVSGVTLSATLSAVKADDFDAQIASODSKINNLTAAQQAAQAVNTIQG 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MKKRILSAVLVSGVTLTGTA--TVNADDFDSKIATDSVINTLSGQAAQAVNTAIKG 58
Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 QVGALESQQSELEAQAQLEAVSQQLGQEIQTLSNKIVARNESLKQQAARSQAQNSAATSY 117
Qy 121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQEQEDKAAVEQKQENQAAINTVAANO 180
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 INTILNSKVSDAINRVSAIREVVSAANEKMLQQEQEDKAAVEQKQENQAAINTVAANO 177
Qy 181 TIAQNTNALNTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 240
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 AIENKKAALATQRAQLEAQAQLELSAQLTTVQNEKASLIQAKAQAEEAARQAQAAEA 237
Qy 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSTE--QSAQAQVN 298
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 KAQAEEAKAQAESVAKAQA---QVESATAPTETVOTQPTTEIKPSNLTATS 286
Qy 299 NSQESTTATEA-----QPSASSASTAA-----VAANTSSANTYPAGQ 336
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
287 SATTVATTATATNEPKVTQPSVVTKAVEAPKAVSVSTPRAVSKPVVRSYDSSNTYPMGQ 346
Qy 337 CTWGVKSLAPWGVNGYWGNGQWAAASAAAGYRGVSTPSAGAVAVW--NDGGYGHVAVVTG 394
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
347 CTWGAKSMAWGVNGYWGNGYWGNGWAGASARAAGYSVGTTPRGAVAVWPDGGYGHVAVVTS 406
Qy 395 V-QGGQIQVQAEANYAGNQSIGNYRGWFPN---GSVSYIYFN 431
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
407 VANNSSIQVMESYAGNMSIGNYRGSPNPSASGSVYIYFN 447

RESULT 8
Q8E2H1_STRAS5 PRELIMINARY; PRT; 447 AA.
AC Q8E2H1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PcsB protein.
GN Name=pcsb; OrderedLocusNames=SAG0017;
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,
RA Wassels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouir H.M., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaudo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an

```

```

RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014192; AAM98925.1; -; Genomic_DNA.
DR TIGR; SAG0017; -.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PSS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 447 AA; 46681 MW; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 58.7%; Score 1235; DB 2; Length 447;
Best Local Similarity 60.1%; Pred. No. 2.7e-47;
Matches 277; Conservative 43; Mismatches 97; Indels 44; Gaps 9;

Qy 1 MKKRILSAVLVSGVTLSATLSAVKADDFDAQIASODSKINNLTAAQQAAQAVNTIQG 60
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MKKRILSAVLVSGVTLTGTA--TVNADDFDSKIATDSVINTLSGQAAQAVNTAIKG 58
Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 120
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
59 QVGALESQQSELEAQAQLEAVSQQLGQEIQTLSNKIVARNESLKQQAARSQAQNSAATSY 117
Qy 121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQEQEDKAAVEQKQENQAAINTVAANO 180
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
118 INTILNSKVSDAINRVSAIREVVSAANEKMLQQEQEDKAAVEQKQENQAAINTVAANO 177
Qy 181 TIAQNTNALNTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNSAATSY 240
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 AIENKKAALATQRAQLEAQAQLELSAQLTTVQNEKASLIQAKAQAEEAARQAQAAEA 237
Qy 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSTE--QSAQAQVN 298
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 KAQAEEAKAQAESVAKAQA---QVESATAPTETVOTQPTTEIKPSNLTATS 286
Qy 299 NSQESTTATEA-----QPSASSASTAA-----VAANTSSANTYPAGQ 336
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
287 SATTVATTATATNEPKVTQPSVVTKAVEAPKAVSVSTPRAVSKPVVRSYDSSNTYPMGQ 346
Qy 337 CTWGVKSLAPWGVNGYWGNGQWAAASAAAGYRGVSTPSAGAVAVW--NDGGYGHVAVVTG 394
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
347 CTWGAKSMAWGVNGYWGNGYWGNGWAGASARAAGYSVGTTPRGAVAVWPDGGYGHVAVVTS 406
Qy 395 V-QGGQIQVQAEANYAGNQSIGNYRGWFPN---GSVSYIYFN 431
Db |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
407 VANNSSIQVMESYAGNMSIGNYRGSPNPSASGSVYIYFN 447

RESULT 9
Q8E7X9_STRAS3 PRELIMINARY; PRT; 447 AA.
AC Q8E7X9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PcsB protein.
GN Name=pcsb; OrderedLocusNames=gsb0016;
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=12154221;
RA Glaser P., Zouine M.C., Buchrieser C., Chevallier F., Frangeul L.,
RA Meadek T., Rousio M., Couve E., Lallouli L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766843; CAD45661.1; -; Genomic_DNA.

```

DR Pfam; PF05257; CHAP; 1.  
DR PRINTS; PR01852; SIBAPROTEIN.  
DR InterPro; IPR009148; Siba.  
DR Pfam; PF05257; CHAP; 1.  
DR PRINTS; PR01852; SIBAPROTEIN.  
DR PROSITE; PS0911; CHAP; 1.  
KW Complete proteome.

SQ SEQUENCE 485 AA; 49165 MW; 8665E4B1FC6CF7CA CRC64;

Query Match 52.0%; Score 1094.5; DB 2; Length 485;  
Best Local Similarity 51.6%; Pred. No. 4.9e-41;  
Matches 252; Conservative 67; Mismatches 108; Indels 61; Gaps 13;

QY 1 MKKRILSAVLVSGVTLSAATLTSAVKADDFDAQIASQDSKINNLTAAQQOAAQAQVNTTIG 60  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 1 MKKRILSAVLVSGVTLSAA--ASVHAEDYDQIAATNNAISNLASIQEEAQAAQAQVATIQS 57  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 61 QVSALQTQAELEQNORLEAQSATLGQIOTLSSKIIVARNESLKQARSQAQKNAAATSY 120  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 58 QVSTLRKTQKEAKNAELEKVSADESEIQELSSKIVARODSLAKQARSQAQNNATSY 117  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 121 INAINSKSVSDAIRNVSAIREVVSANEMKLQOEQDKAAVEQKQENQAANTVAANO 180  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 118 INSILNSKSISEAITRITAISKVVTTANNDDLTKOESDKELAQAQEENQAANTIAANKS 177  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 181 TTAQNTNALNTQAOALEAQLNQELTTAODOKATLVAKAAAEEAARQAAA---AQA 236  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 178 ELETTEAGLTTQOAELEAAQVTLAAELATAQNEKTSLSVSAKSTAESVAASTAASVAOSQA 237  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 237 AAEAARAKALQOA---AQAQAANNNTQA----TDASDOQA--AADNTQAAQTGD 287  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 238 IAESATAQVVASSEATSASSEVAATSEAVAOPSETPVEISTASEAQEPASSETSE 297  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 288 STEQSAQAQVNN-----SDQESTTAT-----EAQPSASSA----- 317  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 298 VQESAAPAVSEAPASVAPVATSEAAPATSEAPASVAPVATSEAPVSEAPVSAAPPVWT 357  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 318 STAAVAANT-----SSANTYPACQCTGWVKSLAPWGVNGWGQWAASAAGAAYRVGS 371  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 358 SEAPAFAETHKVSAASTPTNPVGQCCTGWVKSLAPWAGNMGNNKAKNIASQAAGHSGVT 417  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 372 TPSAGAVAW-ND-GYGHVAVTVGQG-GOIQVOEANYAGNQSIGNRYGFNFP----- 422  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 418 TPVAGIAVWPNDGGYGHVAVTVTSAGANSIQWESNYAGNMSINYRTGTFTDTSSAHG 477  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
QY 423 GSYSYIYP 430  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
DB 478 GSVEFIYP 485

RESULT 11  
QSM6K4 STRT2  
ID OSM6K4\_STRT2 PRELIMINARY; PRT; 474 AA.  
AC OSM6K4;  
DT 01-FEB-2005 (TrEMBLrel. 29, Created)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)  
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
DE Glucan binding protein (PcsB).  
GN Names=pcsb; OrderedLocNames=stu0022;  
OS Streptococcus thermophilus (strain ATCC BAA-250 / LMG 18311).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=264199;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX PubMed=15543133; DOI=10.1038/nbt1034;  
RA Bolotin A., Quinquis B., Renault P., Sorokin A., Ehrlich S.D.,  
RA Kulakauskas S., Lapidus A., Goldstein E., Mazur M., Pusch G.D.,  
RA Fonstein M., Overbeek R., Kyprides N., Furnelle B., Prozzi D.,  
RA Ngui K., Masuy D., Hancy F., Burteau S., Boutry M., Delcours J.,  
RA Goffeau A., Hols P.;  
RT "Complete sequence and comparative genome analysis of the dairy  
bacterium Streptococcus thermophilus";  
RL Nat. Biotechnol. 22:1554-1558(2004).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.



```
RC STRAIN=LMG 18311;
RA Borges F., Layec S., Thibessard A., Fernandez A., Gintz B., Hols P.,
RA Decaris B., Leblond-Bourget N.;
RT "cse, a Chimeric and Variable Gene, Encodes an Extracellular Protein
RT Involved in Cellular Segregation in Streptococcus thermophilus.";
RL J. Bacteriol. 187:2737-2746 (2005).
DR EMBL; CP000023; AAV59752.1; -; Genomic_DNA.
DR EMBL; AY730643; AAW82375.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 474 AA; 48142 MW; 1BACGA9B0A0A200D CRC64;

Query Match 51.9%; Score 1093; DB 2; Length 474;
Best Local Similarity 52.0%; Pred. No. 5.6e-41;
Matches 248; Conservative 67; Mismatches 112; Indels 50; Gaps 12;

QY 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSA---ASVHAEDYDSQIAATNNAISLASQEAQAQVATIQS 57

QY 61 QVSALQTQAELOAENORLEAQSATLQIQITLSSKIVARNESLKOQARSQAQSNATSY 120
DB 58 QVSTLRTQKTELEAKNAELEKVSADLESEIQELSSKIVARQDSLAKQARSQAQNTATSY 117

QY 121 INAININSKVSDAINRVSAIRREVVSANEKMLQOEODKAAVEQKQENQAINTVAANO 180
DB 118 INSILNSKSISEALTRITALSKVVTANNDLITQESDKELAKQENQAINTVAANKS 177

QY 181 TIAQNTNALNTQQAQLAEALQNLQAEELTTAQDQKATLVAQKAAAEAEAAQAAAA 236
DB 178 ELETTEAGLTQQAEEAQQVTLAAELATAGNEKTSILVSAKSTAESVAASTAASVAQSOA 237

QY 237 AAQAKAAAEAKALQEQA---AQQAQAAANNNTQ---TDAQDQAA---AADNTQAAQTGD 287
DB 238 IAESEATAQVVASSEATSVASSEVAATSEAVAQPSVETPVSETSTASEAAQEPASSETSE 297

QY 288 STQSSAAQAVNN-----SDQESTTAT-----EAQPSASSASTAAVAANT 326
DB 298 QPESAAPAVSEAPASVAPVATSEAPATSEAPASVAPVATSEAPASVAPVATSEAPAEATHK 357

QY 327 SSA-----NTYPAGQCTGWGKSLAPWGVNGYNGGQWAAASAAAAGYRVGSTPFSAGAVW- 381
DB 358 VSAASTENTYPVGOCTGWGKSLAPWAGNNGNAKNIASQAAGHSVGTTPVAGATAVWP 417

QY 382 ND-GGYGHVAYVTGVQ-GQIQVEANYAGNQSIGNYRGWFPN-----GSVSYIYP 430
DB 418 NDGGYGHVAYVTGSAGANSIQVMESYAGNMGSINRYGTFTDPTSSAHGGSVFYIYP 474

RESULT 12
Q5XEL1 STRP6 PRELIMINARY; PRT; 398 AA.
AC Q5XEL1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Secreted protein.
GN OrderedLocusNames=M6_Spy0017;
OS Streptococcus pyogenes (serotype M6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301450;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS10394;
RX PubMed=15272401; DOI=10.1086/422697;
RA Banks D.J., Porcella S.F., Barbican K.D., Beres S.B., Phillips L.E.,
RA Voyich J.M., Deleo F.R., Martin J.M., Somerville G.A., Musser J.M.;
RT "Progress toward characterization of the group A Streptococcus
```

```
RT metagenome: complete genome sequence of a macrolide-resistant serotype
RT M6 strain.";
RL J. Infect. Dis. 190:727-738 (2004).
DR EMBL; CP000003; AAT86152.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 51.9%; Score 1092.5; DB 2; Length 398;
Best Local Similarity 53.7%; Pred. No. 5e-41;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSSATTLTSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSSATT---VGAEDLSTKIAKQDSIISLTTTEOKAAQNVSAQ 57

QY 61 QVSALQTQAELOAENORLEAQSATLQIQITLSSKIVARNESLKOQARSQAQSNATSY 120
DB 58 QVSSLQSEQDKLTARNTLEALSQRPEQBIKALTSQIVARNKLEKQARSAYKNNETSGY 117

QY 121 INAININSKVSDAINRVSAIRREVVSANEKMLQOEODKAAVEQKQENQAINTVAANO 180
DB 118 INALLNSKISIDVNVNLVAINRAVSAKALLEQOKADKVSLEEKQAANTQAITIANMA 177

QY 181 TIAQNTNALNTQQAQLAEALQNLQAEELTTAQDQKATLVAQKAAAEAEAAQAAAA 240
DB 178 MAEQNTLRTQANLEAATANLALQASATEDKANLVAQKEAAEAKAAEALAEQQA 237

QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSTEQSAAQAVNNS 300
DB 238 KAQEQA---AQQAASVEAAKSAITPAQA-----TPAAQSSNAIEPAALTA--- 280

QY 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWGVNGYNGGQWAA 360
DB 281 -----PAAPSARPOTS-----YDSNTYPVGOCTGWGKSLAPWAGNNGGQW 326

QY 361 SAAAAGYRVGSTPFSAGAVAWNDGGYGHVAYVTGVQ-GQIQVEANYAGNQSIGNYRGW 419
DB 327 SAQAAGYRVGSTPFWGAIWVDGGYGHVAVVVEVSASSIRVMESYSGRQIADHRGW 386

QY 420 FNPGSYIYPN 431
DB 387 ENPTGVTFIYPH 398

RESULT 13
Q7CNQ7 STRP8 PRELIMINARY; PRT; 398 AA.
AC Q7CNQ7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=spvM18_0020;
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301451;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS8232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108; DOI=10.1073/pnas.062526099;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.W., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.E., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";
```

```

RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
DR EMBL; AE009955; AAL96849.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 51.9%; Score 1092.5; DB 2; Length 398;
Best Local Similarity 53.7%; Pred. No. 5e-41;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQQAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQQAQAQVNTIQG 60
QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQAQVNTIQG 120
DB 61 QVSALQTOQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQAQVNTIQG 120
QY 121 INALINKSVSDAINRVSAIREVVSAANEKMLQOEQDKAAVEQKQAEQAQVNTIQG 180
DB 121 INALINKSVSDAINRVSAIREVVSAANEKMLQOEQDKAAVEQKQAEQAQVNTIQG 180
QY 181 TIAQNTNALNTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQAQVNTIQG 240
DB 181 TIAQNTNALNTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQAQVNTIQG 240
QY 241 KAAAEKALQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQAQVNTIQG 300
DB 241 KAAAEKALQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQAQVNTIQG 300
QY 301 QDSTTTATEAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWGVNMGNGGOWAY 360
DB 301 QDSTTTATEAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWGVNMGNGGOWAY 360
QY 361 SAAAGYRVGSGTSPAGAVAVWNDGGYGHVAVYTGVO-GGQIQVEANYAGNQSIGNRGW 419
DB 361 SAAAGYRVGSGTSPAGAVAVWNDGGYGHVAVYTGVO-GGQIQVEANYAGNQSIGNRGW 419
QY 420 FNPQSVSYIYPN 431
DB 420 FNPQSVSYIYPN 431
QY 387 FNPQSVSYIYPN 398
DB 387 FNPQSVSYIYPN 398

RESULT 15
Q9A128_STRPY PRELIMINARY; PRT; 398 AA.
AC Q9A128; Q7BH59;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Secreted protein Siba precursor.
GN OrderedLocusNames=SPY0019;
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;

Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
DR EMBL; AE009955; AAL96849.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 51.9%; Score 1092.5; DB 2; Length 398;
Best Local Similarity 53.7%; Pred. No. 5e-41;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQQAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQQAQAQVNTIQG 60
QY 61 QVSALQTOQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQAQVNTIQG 120
DB 61 QVSALQTOQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQAQVNTIQG 120
QY 121 INALINKSVSDAINRVSAIREVVSAANEKMLQOEQDKAAVEQKQAEQAQVNTIQG 180
DB 121 INALINKSVSDAINRVSAIREVVSAANEKMLQOEQDKAAVEQKQAEQAQVNTIQG 180
QY 181 TIAQNTNALNTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQAQVNTIQG 240
DB 181 TIAQNTNALNTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQAQVNTIQG 240
QY 241 KAAAEKALQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQAQVNTIQG 300
DB 241 KAAAEKALQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAQAQVNTIQG 300
QY 301 QDSTTTATEAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWGVNMGNGGOWAY 360
DB 301 QDSTTTATEAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWGVNMGNGGOWAY 360
QY 361 SAAAGYRVGSGTSPAGAVAVWNDGGYGHVAVYTGVO-GGQIQVEANYAGNQSIGNRGW 419
DB 361 SAAAGYRVGSGTSPAGAVAVWNDGGYGHVAVYTGVO-GGQIQVEANYAGNQSIGNRGW 419
QY 420 FNPQSVSYIYPN 431
DB 420 FNPQSVSYIYPN 431
QY 387 FNPQSVSYIYPN 398
DB 387 FNPQSVSYIYPN 398

RESULT 14
Q9P318_STRP3 PRELIMINARY; PRT; 398 AA.
AC Q9P318; Q7CFL7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=SPS0015; SpvM3 0014;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301448;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yasunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution.";
RL Genome Res. 13:1042-1055(2003).
RN [2]
RP NUCLEOTIDE SEQUENCE.

```

RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,  
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";  
RN Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RA Fagan P.K., Reinscheid D., Gottschalk B., Chhatwal G.S.;  
RT "Identification and characterization of a novel secreted protein from  
RT group A streptococcus";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE006474; AAK33158.1; -; Genomic\_DNA.  
DR EMBL; AF319999; AAL73135.1; -; Genomic\_DNA.  
DR InterPro; IPR007921; CHAP.  
DR InterPro; IPR009148; Siba.  
DR Pfam; PF05257; CHAP; 1.  
DR PRINTS; PR01852; SIBAPROTEIN.  
DR PROSITE; PS00911; CHAP; 1.  
DR Complete proteome; Signal.  
FT SIGNAL 1 23  
SQ SEQUENCE 398 AA; 41899 MW; 28A9B3F7195E969B CRC64;  
  
Query Match 51.6%; Score 1085.5; DB 2; Length 398;  
Best Local Similarity 53.0%; Pred. No. le-40;  
Matches 229; Conservative 60; Mismatches 108; Indels 35; Gaps 5;  
  
Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASODSKINNLTACQAAQAVNTIQG 60  
Db 1 MKKRILSAVLVSGVTLSAATT---VGAEDLSTKIAKQDSIISNLTTEQKAAQNVSAQA 57  
  
Qy 61 QVSALOTQQAELQAEQORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSAAKSNAAATSY 120  
Db 58 QVSSLSQSEQDKLTARNTLEALSRRFEQIKALTSQIVARNEKLKNQARSAYKNNETSGY 117  
  
Qy 121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQEQDKAAVEQKQEQENQAAINTVAANQE 180  
Db 118 INALLNSKISDVVNRLVAINRAVSAKAKLLEQKADKVSLEEKQAAANQTAINTIAANMA 177  
  
Qy 181 TIAQNTNALNTQQAQLEAAQNLQAEILTTAQDOKATLVAOKAAAEAAQAAAAQAAEA 240  
Db 178 MAEENQNTLTQQAANLVAAANTANLALQASATEDKANLVAKAEAKAAEALAEQAAKV 237  
  
Qy 241 KAAAEAKALQEQAAQAAANNTQATDASDQAAAAADNTQAAQTDGDSQSAQAQVNN 300  
Db 238 KAQEQA---AQQAASVEAKSAITPAPQA-----TPAQSSNAIEPAL----- 278  
  
Qy 301 DQESTTATEQPSASSASTAAVAANTSSANTYPAGQCTGWVKSLAPWVGNWNGGQWAA 360  
Db 279 -----TAPAAPSAQPSQTSYDSSNTYPVGQCTGWGAKSLAPWAGNNWNGGQWAY 326  
  
Qy 361 SAAAGYRVGSGTSAGAVAVWNDGGYGHVAVVTQVQ--GGIOIYOAEANYAGNOSIGNYRGW 419  
Db 327 SAQAAGYRTGSGTSMVGAIAVWNDGGYGHVAVVEVQSASSIRVNESNYSGRYIADHRGW 386  
  
Qy 420 FNGSGVSYIYPN 431  
Db 387 FNTGTFTIYPH 398

Search completed: February 10, 2006, 22:31:19  
Job time : 296.878 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:31:39 ; Search time 77.7712 Seconds  
(without alignments)  
458.180 Million cell updates/sec

Title: US-10-797-821-30

Perfect score: 2104

Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNYRGWFPNGSVSYIPN 431

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1057.5	50.3	399	2	US-09-107-433-3230
2	1056.5	50.2	392	2	US-09-583-110-4374
3	674	32.0	461	1	US-08-186-222-2
4	590	28.0	525	2	US-09-107-532A-5095
5	562	26.7	210	2	US-09-222-938A-67
6	528.5	25.1	497	2	US-09-134-000C-5990
7	503	23.9	449	2	US-09-071-035-482
8	503	23.9	449	2	US-10-206-576-482
9	503	23.9	450	2	US-09-134-000C-5714
10	454	21.6	422	2	US-09-071-035-484
11	454	21.6	422	2	US-10-206-576-484
12	279	13.3	469	2	US-09-489-039A-13565
13	236.5	11.2	2310	2	US-09-874-923-120
14	233	11.1	257	2	US-09-710-279-3244
15	233	11.1	264	2	US-09-134-001C-5035
16	233	11.1	267	2	US-09-134-001C-4539
17	230.5	11.0	477	2	US-09-902-540-11649
18	224.5	10.7	1236	2	US-09-769-787-109
19	223	10.6	468	2	US-09-328-352-6321
20	215.5	10.2	610	2	US-09-336-447A-11
21	215.5	10.2	610	2	US-09-952-267B-11
22	215	10.2	149	2	US-09-710-279-1682
23	215	10.2	157	2	US-09-710-279-2870
24	213	10.1	270	2	US-09-134-001C-5441
25	213	10.1	639	2	US-09-902-540-14908
26	211.5	10.1	440	2	US-08-302-756B-35
27	209	9.9	266	2	US-09-134-001C-5453

28	209	9.9	1070	2	US-09-902-540-13861	Sequence 13861, A
29	208	9.9	679	2	US-09-489-039A-12307	Sequence 12307, A
30	206.5	9.8	1566	1	US-08-687-956A-23	Sequence 23, Appl
31	200.5	9.5	955	1	US-08-006-676B-1	Sequence 1, Appli
32	200.5	9.5	955	1	US-08-282-845-2	Sequence 2, Appli
33	200.5	9.5	955	1	US-08-428-414A-3	Sequence 3, Appli
34	200.5	9.5	955	4	PCT-US94-00324-1	Sequence 1, Appli
35	199	9.5	655	2	US-09-902-540-10005	Sequence 10005, A
36	198	9.4	1024	2	US-09-270-767-44973	Sequence 23, Appl
37	198	9.4	1561	2	US-08-894-017-23	Sequence 44973, A
38	198	9.4	1561	2	US-09-456-474-23	Sequence 23, Appl
39	198	9.4	1565	6	5352450-2	Patent No. 5352450
40	195	9.3	1095	2	US-09-710-279-3154	Sequence 3154, Ap
41	194.5	9.2	2101	1	US-08-466-390-4	Sequence 4, Appli
42	194.5	9.2	2101	1	US-08-470-950-4	Sequence 4, Appli
43	194.5	9.2	2101	1	US-08-467-781-4	Sequence 4, Appli
44	194.5	9.2	2101	1	US-08-195-487-4	Sequence 4, Appli
45	194.5	9.2	2101	1	US-08-483-924-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1  
US-09-107-433-3230  
; Sequence 3230, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; THERAPEUTICS

NUMBER OF SEQUENCES: 5206

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

MEDIUM TYPE: CD/ROM ISO9660

COMPUTER: <Unknown>

OPERATING SYSTEM: <Unknown>

SOFTWARE: <Unknown>

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,433

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/ 085131

FILING DATE: May 12, 1998

APPLICATION NUMBER: 60/051553

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-011

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 3230:

SEQUENCE CHARACTERISTICS:

LENGTH: 399 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: YES

ORIGINAL SOURCE:

ORGANISM: Streptococcus pneumoniae

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (B) LOCATION 1...399

SEQUENCE DESCRIPTION: SEQ ID NO: 3230:

US-09-107-433-3230

Query Match 50.3%; Score 1057.5; DB 2; Length 399;  
Best Local Similarity 50.3%; Pred. No. 5.1e-69;  
Matches 219; Conservative 73; Mismatches 92; Indels 51; Gaps 5;

QY 1 MKKRILSAVLVSGVTLSSTATLSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60  
DB 8 MKKKILASLLSTVMVSQVAVLTAAHTTDDKIAQDNKISNLTAAQQAAQAVNTIQG 67

QY 61 QVSALQTQQAELQAEORLEAQSATLGGQITQSSKIVARNESLKKQARSAQSKNAATSY 120  
DB 68 QVSAIQAEQSNLQAEORLEAQSATLGGQITQSSKIVARNESLKKQARSAQSKNAATSY 127

QY 121 INALINSKSVDAINRVSAIREVVUSANEKMLQOEODKAAVEKQOQENQAAINTVAANO 180  
DB 128 INTIVNSKSITEAISRVAAWSEIVSANNKMLEQOORADKKAISEKQVANNDAINTVIAN 187

QY 181 TIAQNTNALNTQQAELQAEORLEAQSATLGGQITQSSKIVARNESLKKQARSAQSKNA 240  
DB 188 KLADDAQAULTTKQAELKAAELSLAAEKATAEGEKASILLEKQAAAEAEARAAVAEAYK- 246

QY 241 KAAAEAKALQEAQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSSTEQSAAQAV 300  
DB 247 -----EKRAQQOQSVLASANTNLTAQVAVSESAAA----- 277

QY 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGCTGKVSILAPWGVNMGNGOWAA 360  
DB 278 -----PVRKVRPTYS-----TNASSYPIGECTGKVTLPWAGDYGNGAQWAT 322

QY 361 SAAAAGRYVSGTSPSAGAVAVMNDGGYGHVAVYVTVGQG-GQIQVQEAANYAGNQSIGNY 419  
DB 323 SAAAAGRTGSTPQVGAICWMDGGYGHVAVVTVAVESTTTRIQVSESNYAGNRTIGNH 382

QY 420 FNP-----GSVSYIY 429  
DB 383 FNPPTTSEGFTVIY 397

RESULT 2

US-09-583-110-4374  
; Sequence 4374, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 4374  
; LENGTH: 392  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-4374

Query Match 50.2%; Score 1056.5; DB 2; Length 392;  
Best Local Similarity 50.3%; Pred. No. 5.9e-69;  
Matches 219; Conservative 72; Mismatches 93; Indels 51; Gaps 5;

QY 1 MKKRILSAVLVSGVTLSSTATLSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60  
DB 1 MKKKILASLLSTVMVSQVAVLTAAHTTDDKIAQDNKISNLTAAQQAAQAVNTIQG 60

QY 61 QVSALQTQQAELQAEORLEAQSATLGGQITQSSKIVARNESLKKQARSAQSKNAATSY 120

Db 61 QVSAIQAEQSNLQAEORLEAQSATLGGQITQSSKIVARNESLKKQARSAQSKNAATSY 120

QY 121 INALINSKSVDAINRVSAIREVVUSANEKMLQOEODKAAVEKQOQENQAAINTVAANO 180

Db 121 INTIVNSKSITEAISRVAAWSEIVSANNKMLEQOORADKKAISEKQVANNDAINTVIAN 180

QY 181 TIAQNTNALNTQQAELQAEORLEAQSATLGGQITQSSKIVARNESLKKQARSAQSKNA 240

Db 181 KLADDAQAULTTKQAELKAAELSLAAEKATAEGEKASILLEKQAAAEAEARAAVAEAYK- 239

QY 241 KAAAEAKALQEAQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSSTEQSAAQAV 300

Db 240 -----EKRAQQOQSVLASANTNLTAQVAVSESAAA----- 270

QY 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGCTGKVSILAPWGVNMGNGOWAA 360

Db 271 -----PVRKVRPTYS-----TNASSYPIGECTGKVTLPWAGDYGNGAQWAT 315

QY 361 SAAAAGRYVSGTSPSAGAVAVMNDGGYGHVAVYVTVGQG-GQIQVQEAANYAGNQSIGNY 419

Db 316 SAAAAGRTGSTPQVGAICWMDGGYGHVAVVTVAVESTTTRIQVSESNYAGNRTIGNH 375

QY 420 FNP-----GSVSYIY 429

Db 376 FNPPTTSEGFTVIY 390

RESULT 3

US-08-186-222-2  
; Sequence 2, Application US/08186222  
; Patent No. 5559007  
; GENERAL INFORMATION:  
; APPLICANT: Suri, Bruno  
; TITLE OF INVENTION: Bacterial Vectors  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CIBA-GEIGY Corporation  
; STREET: 7 Skyline Drive  
; CITY: Hawthorne  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10532  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/186,222  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/672,205  
; FILING DATE: 19-MAR-1991  
; APPLICATION NUMBER: GB 9006400.7  
; FILING DATE: 22-MAR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Villamizar, JoAnn  
; REGISTRATION NUMBER: 30,598  
; REFERENCE/DOCKET NUMBER: 4-17994/A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (914)785-7121  
; TELEFAX: (914)347-5769  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 461 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-186-222-2

Query Match 32.0%; Score 674; DB 1; Length 461;



```
; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-67

Query Match      26.7%; Score 562; DB 2; Length 210;
Best Local Similarity 55.2%; Pred. No. 2.2e-33;
Matches 116; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASODSKINNLTAAQQAAQAQVNTIOG 60
Db 1 MKKKILASLLSTVWVSQVAVLTTHAETTTDDKIAAQDNKISNLTAAQQEAQKQVDIOE 60

Qy 61 QVSALOTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
Db 61 QVSAIQAEQSNLQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120

Qy 121 INAIINSKVSDAINRVSAITREVVSANERKMLQOEQDKAAVEQKQAEQQAINTVAANOE 180
Db 121 INTVNSKISITEAISRVAAAMEIVSANNKMLEQKADKKAISEKQVANNDAINTVIANQ 180

Qy 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTA 210
Db 181 KLADDAQALTTKQELKAAELSLAAEKATS 210

RESULT 6
US-09-134-000C-5990
; Sequence 5990, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5990
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5990

Query Match      25.1%; Score 528.5; DB 2; Length 497;
Best Local Similarity 31.4%; Pred. No. 1.8e-30;
Matches 152; Conservative 82; Mismatches 175; Indels 75; Gaps 12;

Qy 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASODSKINNLTAAQQAAQAQVNTIOG 60
Db 16 LKKSLSALMVCSTILTSVALPSPANFADEYDTKIQQQDKINALTSQMSDAEKVAIEN 75

Qy 61 QVSALOTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
Db 76 DMVETAKQIDTLTAKKNLSSEVSKLYSEISDLNVRILQREVQMTKQARDVQVNGQSDSI 135

Qy 121 INAIINSKVSDAINRVSAITREVVSANERKMLQOEQDKAAVEQKQAEQQAINTVAANOE 180
Db 136 IDAVLDADSVADAIGRVQAVSTWMSANNLEQEKDKATVEKTKNVEKQIAELEAATK 195

Qy 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAQQDQKATLV-----AKAAAEAAARQAAAA- 234
Db 196 ELNDKTESLTKLIQAEVANDLEAQRSSEQGKRGDFIKKQAEKGLAEQARQRAAK 255

Qy 235 -----QAAAEAKAAAEAKALQEAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSTE 290
Db 256 KAEQEAQAQAQAQAQ-KAAAEQAQATKAA---NEAASAAAEKAA----TPVVSSTTTE 307

Qy 291 QSAQAQVNSDDQS-----TTATAQP-----SASSASTAA 321
Db 308 STTTOETTSSTETESVVTTFVAAAPAEKEKEVPVNTPTTPEKGNKAPGNGVTSQKQA 367

; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-67

Query Match      26.7%; Score 562; DB 2; Length 210;
Best Local Similarity 55.2%; Pred. No. 2.2e-33;
Matches 116; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASODSKINNLTAAQQAAQAQVNTIOG 60
Db 1 MKKKILASLLSTVWVSQVAVLTTHAETTTDDKIAAQDNKISNLTAAQQEAQKQVDIOE 60

Qy 61 QVSALOTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
Db 61 QVSAIQAEQSNLQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120

Qy 121 INAIINSKVSDAINRVSAITREVVSANERKMLQOEQDKAAVEQKQAEQQAINTVAANOE 180
Db 121 INTVNSKISITEAISRVAAAMEIVSANNKMLEQKADKKAISEKQVANNDAINTVIANQ 180

Qy 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTA 210
Db 181 KLADDAQALTTKQELKAAELSLAAEKATS 210

RESULT 7
US-09-071-035-482
; Sequence 482, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 482:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-071-035-482

Query Match      23.9%; Score 503; DB 2; Length 449;
Best Local Similarity 31.4%; Pred. No. 1.1e-28;
Matches 141; Conservative 88; Mismatches 176; Indels 44; Gaps 12;

Qy 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASODSKINNLTAAQQAAQAQVNTIOG 60
Db 1 VKKELFASVLLCSLTLSAIAITPSIALADNVDKIEKNQEISSLKAKQGLAQSVSSLEA 60

Qy 61 QVSALOTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120
Db 61 EVSSVFDESMALEQKQTLKAKSEQLQEQETLNLNRIEKNRNEAIKNQARDVQVNGOSTTM 120

Qy 121 INAIINSKVSDAINRVSAITREVVSANERKMLQOEQDKAAVEQKQAEQQAINTVAANOE 180
Db 121 LDVLDADSVADAIGRVQAVSTWMSANNLEQEKDKATVEKTKNVEKQIAELEAATK 180

Qy 181 TIAQNTNALNTQQAQLEAAQLNLQAEELTTAQQDQKATLVAAQKAAAE-EAARQAAAAQAAAE 239
```



Db 181 ELETRODILSKOSELVNVMKASLALEQSSAESKAGLEKQKAAAEQARLAAEQKAAAE 240  
Qy 240 -AKAAAEAKALQEQAAQAAANNTQ-----ATDASDQAAA-----AADNTQAAQ 284  
Db 241 KAKQAAAKPAKAEVKAEPVASSSTTEAQAPASSSSATESSTQTTTTPSTDNSATEN 300  
Qy 285 TGDSTEQSAQAV-----NNSDOESTTATEAQPSASSASTAAVAANTSSANTYPAGCQT 338  
Db 301 TGSSESQPVQPTTPSDNGNGGQTGGTVPPTPEPTPAPSADPTINALNLVRLQSLG--- 357  
Qy 339 WGVKSLAPWYGNVWNGGOWAASAAAAGYRVGSPSAGAVAVWMDGGYGHVAVYTVGGG 398  
Db 358 -----LRPVV---WDAGLAASATARAQVEAGGIPNDH-----WSRG---DEVIALMWAPGN 403  
Qy 399 QIOV---QEANYAGNOSIGNYRGW-FNPG 423  
Db 404 SVMAMWYNETNMVTASGSG-HRDWEINPG 431

RESULT 8  
US-10-206-576-482  
; Sequence 482, Application US/10206576  
; Patent No. 6913907  
; GENERAL INFORMATION:  
; APPLICANT: Choi et al.  
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides  
; NUMBER OF SEQUENCES: 497  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Human Genome Sciences, Inc.  
; STREET: 9410 Key West Avenue  
; CITY: Rockville  
; STATE: Maryland  
; COUNTRY: USA  
; ZIP: 20850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-R  
; COMPUTER: Dell Latitude  
; OPERATING SYSTEM: Windows 98  
; SOFTWARE: ASCII Text  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/206,576  
; FILING DATE: 29-Jul-2002  
; CLASSIFICATION: <unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/071,035  
; FILING DATE: 1998-05-04  
; APPLICATION NUMBER: US 60/046,655  
; FILING DATE: 1997-03-16  
; APPLICATION NUMBER: US 60/044,031  
; FILING DATE: 1997-05-06  
; APPLICATION NUMBER: US 60/066,009  
; FILING DATE: 1997-11-14  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Hyman, Mark J.  
; REGISTRATION NUMBER: 46,789  
; REFERENCE/DOCKET NUMBER: PB369P1D1  
; INFORMATION FOR SEQ ID NO: 482:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 449 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; SEQUENCE DESCRIPTION: SEQ ID NO: 482:  
US-10-206-576-482

Query Match 23.9%; Score 503; DB 2; Length 449;  
Best Local Similarity 31.4%; Pred. No. 1.1e-28;  
Matches 141; Conservative 88; Mismatches 176; Indels 44; Gaps 12;  
Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDDFAQIASQDSKINNLTAAQQAAQAVNTIQG 60

Db 1 VKKRLPASVLLCSLTLSAATPSIALADNVDKKIEKNQEISSLKAKQGDLSQVSSLEA 60  
Qy 61 QVSALOTQQAELQAEQNRLEQAQATLGGQQTQTLSSKIVARNESLKQARSAQKSNATSY 120  
Db 61 EVSSVFDESALREKQKTLKAKSEQLQEQEITNLNRIEKNRATKNOARDVQVNGQSTTM 120  
Qy 121 INAIINSKVSVDAINRVSAIREVVVSANEKMLQQEQODKAAVEKQEQENQAINTVAANO 180  
Db 121 LDAVLDAQSVADAIISRQAVSTIVSANNLQMQQEKDQAVVDKAEENEKKVKOLEATEA 180  
Qy 181 TIAQNTNALNTQQAELQAEQNLQAEITTAQDQKATLVAQKAAAE-EAARQAAQAAAE 239  
Db 181 ELETRODILSKOSELVNVMKASLALEQSSAESKAGLEKQKAAAEQARLAAEQKAAAE 240  
Qy 240 -AKAAAEAKALQEQAAQAAANNTQ-----ATDASDQAAA-----AADNTQAAQ 284  
Db 241 KAKQAAAKPAKAEVKAEPVASSSTTEAQAPASSSSATESSTQTTTTPSTDNSATEN 300  
Qy 285 TGDSTEQSAQAV-----NNSDOESTTATEAQPSASSASTAAVAANTSSANTYPAGCQT 338  
Db 301 TGSSESQPVQPTTPSDNGNGGQTGGTVPPTPEPTPAPSADPTINALNLVRLQSLG--- 357  
Qy 339 WGVKSLAPWYGNVWNGGOWAASAAAAGYRVGSPSAGAVAVWMDGGYGHVAVYTVGGG 398  
Db 358 -----LRPVV---WDAGLAASATARAQVEAGGIPNDH-----WSRG---DEVIALMWAPGN 403  
Qy 399 QIOV---QEANYAGNOSIGNYRGW-FNPG 423  
Db 404 SVMAMWYNETNMVTASGSG-HRDWEINPG 431

RESULT 9  
US-09-134-000C-5714  
; Sequence 5714, Application US/09134000C  
; Patent No. 6617156  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; FILE REFERENCE: 032796-032  
; CURRENT APPLICATION NUMBER: US/09/134,000C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR FILING DATE: 1997-08-15  
; NUMBER OF SEQ ID NOS: 6812  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 5714  
; LENGTH: 450  
; TYPE: PRT  
; ORGANISM: Enterococcus faecalis  
US-09-134-000C-5714

Query Match 23.9%; Score 503; DB 2; Length 450;  
Best Local Similarity 31.4%; Pred. No. 1.1e-28;  
Matches 141; Conservative 88; Mismatches 176; Indels 44; Gaps 12;  
Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDDFAQIASQDSKINNLTAAQQAAQAVNTIQG 60  
Db 2 VKKRLPASVLLCSLTLSAATPSIALADNVDKKIEKNQEISSLKAKQGDLSQVSSLEA 61  
Qy 61 QVSALOTQQAELQAEQNRLEQAQATLGGQQTQTLSSKIVARNESLKQARSAQKSNATSY 120  
Db 62 EVSSVFDESALREKQKTLKAKSEQLQEQEITNLNRIEKNRATKNOARDVQVNGQSTTM 121  
Qy 121 INAIINSKVSVDAINRVSAIREVVVSANEKMLQQEQODKAAVEKQEQENQAINTVAANO 180  
Db 122 LDAVLDAQSVADAIISRQAVSTIVSANNLQMQQEKDQAVVDKAEENEKKVKOLEATEA 181  
Qy 181 TIAQNTNALNTQQAELQAEQNLQAEITTAQDQKATLVAQKAAAE-EAARQAAQAAAE 239  
Db 182 ELETRODILSKOSELVNVMKASLALEQSSAESKAGLEKQKAAAEQARLAAEQKAAAE 241  
Qy 240 -AKAAAEAKALQEQAAQAAANNTQ-----ATDASDQAAA-----AADNTQAAQ 284



```
Qy 28 DFDQAQASQDSKINNLTAAQQAQAQVNTIQGVVSALQTCQAELQAEHQENRLEBAQSATLG 87
Db 1 DNDVKKTEENQETSSSLKAKQGDLSAQVSSLEAEVSVFDESMALEKQKOTLAKSBLQ 60
Qy 88 QOJOTLSSKIVARNESLKQQAARSQAQKNAATSYINALINSKVSVDATNRVSAIREVVSA 147
Db 61 QETITNLMORIEKNEATKQARDVQVNGQSTTMDLDAVIDADSDVADALSRVQAVSTIVSAN 120
Qy 148 EKMLQQEQEDKAAVEQKQENQAAINTVAANQETIAQNTNALINTQQAQLEAAQLNLQAE 207
Db 121 NDLMQQKEDKQAVVDKKAENKVKQLEATEALETKRQDILLKQSELNVKMSLALEQ 180
Qy 208 TTQADOKATLVAQAAAE-BAARQAAAAQAAAE-AKAAAEAKALQEQAAQAAANNTQ 265
Db 181 SSAESSKAGLEKQAAAEQARLAAEQKAAAEKAKQAAAKPAKAEVKAEPVASSSTTE 240
Qy 266 -----ATDASDQAAA-----AADNTQAOTGDSTEQSAQAQV-----NNSQEST 305
Db 241 AQAPASSSSATESSTQTTTTPSTDNSTATENTGSSSSSQPVQPTTPSDNGNNGGTGG 300
Qy 306 TATEAOPSASSASTAAVAANTSSANTYPAGQCTMGVKS LAPWVGNVNGGQWAAASAAAA 365
Db 301 GTVTPPTPEPTAPSADPTINALNVLRSLG-----LRPV-----WDAGLAASATARAA 349
Qy 366 GYRVGTPSAGAVAVMNDGQYGHVAVYTVGVGGQIQV---QEANYAGNQSIGNYRGH-FN 421
Db 350 QVEAGGIPNDH---WSRG--DEVIAIMWAPGNSVIMAWYNETNMVMTASGSG-HRDWEIN 402
Qy 422 PG 423
Db 403 PG 404

RESULT 12
US-09-489-039A-13565
; Sequence 13565, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 13565
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-13565

Query Match 13.3%; Score 279; DB 2; Length 469;
Best Local Similarity 29.0%; Pred. No. 2.2e-12;
Matches 113; Conservative 59; Mismatches 163; Indels 56; Gaps 12;

Qy 31 DAQIASQDSKINNLTAAQQAQVNTIQGVVSALQTCQAELQAEHQENR 78
Db 84 DAIMVDPGAVVNNYNNRQOQQAASARRAAEQKAAQOQAAELREKQAAEQERLKLQEQR 143
Qy 79 LEAQSATLGOQIOTLSSKIVARNESLKQQAARSQAQKNAATSYINALINSKVSVDATNRVS 138
Db 144 LQAEAAKEAQEQ-----QKQAEAAAKAAAKAAKAKADQAKAEQAEAAKAA 190
Qy 139 AIREVVSANEMKMQQEQSDKAAVEQKQENQAAINTVA-ANQETIAQNTNALINTQQAQLE 197
Db 191 AEAKAKADAQAKAAEQAAAKAAADAKQAAEAAAKAAAKAAAEAKQAAEAAKAAAEQAQKAE 250
Qy 198 AA---QLNLQAEITTAQDOKATLVAQKAAAEAAARQAAAAQAAAEKAAAEKALQEQAA 254
Db 251 AAAAKAQAEKAAQEQAAKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAAAEKAA 309
```

```
Qy 255 QAQAANNTQATDASDQQAQAADNTQAAQTGD-----STEQSAAQAQVNNSDQEST 305
Db 310 AAEKAAAD--KAAKAAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 367
Qy 306 TATEAOPSASSASTA-----AAVANTSSANTYPAGQCTMGVKS LAPWVGNVNGGQWAAASAAAA 352
Db 368 AAAGSGNTRKNSAGADINNYAGQIKSAIESKFYDASSYAGKTCITLRIK-LAPDGLLLNIQ 426
Qy 353 GNGGQWA-ASAAAAGYRVGSTPSAGAVAV 381
Db 427 SEGDPALCOAALAAARQAKFPKPPSOAVY 456

RESULT 13
US-09-874-923-120
; Sequence 120, Application US/09874923
; Patent No. 6638517
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Webb, John R.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Bhatia, Ajay
; APPLICANT: Coler, Rhea
; APPLICANT: Probat, Peter
; APPLICANT: Brannon, Mark
; TITLE OF INVENTION: LEISHMANIA ANTIGENS FOR USE IN THE
; FILE REFERENCE: 210121.420C8
; CURRENT APPLICATION NUMBER: US/09/874,923
; CURRENT FILING DATE: 2001-06-04
; NUMBER OF SEQ ID NOS: 122
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 120
; LENGTH: 2310
; TYPE: PRT
; ORGANISM: Leishmania major and chagasi
US-09-874-923-120

Query Match 11.2%; Score 236.5; DB 2; Length 2310;
Best Local Similarity 28.4%; Pred. No. 2.1e-08;
Matches 97; Conservative 51; Mismatches 140; Indels 53; Gaps 12;

Qy 26 KADDFDAQASQDSKINNLTAAQQAQAQV-----NITQGVVSALQTCQAELQAEHQENRLEA 81
Db 1049 RLDTATQQAELARVARLAADREARQQLAANAELQORLDTATQQAELARVARLAA 1108
Qy 82 QSATLGOQIOTLS-----SKIVARNESLKQQAARSQAQKNAATSYINALINSKVS 131
Db 1109 NAEELQORLDTATQQAELARVARLAADREARQQLAANAELQORLDTATQQAELARVAR 1161
Qy 132 DAINRVSATREVVSANEMKMQQEQEDKAAVEQKQENQAAINTVAANQETIAQNTNALNT 191
Db 1162 LAADRDEA-RQQLAANAELQOR-LDTAT--QQAELARVARLAANAELQORLDTATQ 1217
Qy 192 QQAQLEA-AQL-----NLQAEITTAQDOKATLVAQKA-----AAEAAA 228
Db 1218 QRALEAQLARLAADREARQQLAANAELQORLDTATQQAELARVARLAANAELQ 1277
Qy 229 RQAAAAQAAAEKAAAEKALQEQAAQAAQAAANNN--TQATDASDQQAQAADNTQAA--AQ 284
Db 1278 RLDTATQQAELARVARLAADREARQQLAANAELQORLDTATQQAELARVARLAA 1337
Qy 285 TGDSTEQSAAQVNNSDQESTTATEAOPSASSASTAAVAAN 325
Db 1338 DGDEARQQLAANAELQORLDTATQ-QRAELARVARLAAN 1377

RESULT 14
US-09-710-279-3244
; Sequence 3244, Application US/09710279
```

Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: P0348008  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3244  
; LENGTH: 257  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-09-710-279-3244

Query Match 11.1%; Score 233; DB 2; Length 257;  
Best Local Similarity 34.5%; Pred. No. 2.2e-09;  
Matches 59; Conservative 29; Mismatches 75; Indels 8; Gaps 5;  
  
QY 261 NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNSDQESTTATEAQPSSASTA 320  
Db 91 NNYSNYNNYNYQ---SNNTQSQRTTPTGGLGASYSTSSNVHVTTTSA-PSSNGVSL 146  
QY 321 AVAANTSSANTYPAGQCTGCV-KSLAPWGVNGVNGGQWAAASAAAAGYRVGSTPSAGAVA 379  
Db 147 --NARGASGNLYTSGQCTYYVDFRVGKGIGSTWGNANNNWANAARSGYTVNNSPAKGAIL 204  
QY 380 VVNDGGYGHVAYVTGV-QGGQIQVEANYAGNOSIGNYRCWFNPGSVSYIY 429  
Db 205 QTSQGAYGHVAYVEGVNSNGSIRVSEMYGHGAGVVTSTRTISASQAASNY 255

RESULT 15  
US-09-134-001C-5035  
; Sequence 5035, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC-007  
; CURRENT APPLICATION NUMBER: US/09/134,001C  
; CURRENT FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: US 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: US 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 5674  
; SEQ ID NO 5035  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-5035

Query Match 11.1%; Score 233; DB 2; Length 264;  
Best Local Similarity 34.5%; Pred. No. 2.3e-09;  
Matches 59; Conservative 29; Mismatches 75; Indels 8; Gaps 5;  
  
QY 261 NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNSDQESTTATEAQPSSASTA 320  
Db 98 NNYSNYNNYNYQ---SNNTQSQRTTPTGGLGASYSTSSNVHVTTTSA-PSSNGVSL 153  
QY 321 AVAANTSSANTYPAGQCTGCV-KSLAPWGVNGVNGGQWAAASAAAAGYRVGSTPSAGAVA 379  
Db 154 --NARGASGNLYTSGQCTYYVDFRVGKGIGSTWGNANNNWANAARSGYTVNNSPAKGAIL 211  
QY 380 VVNDGGYGHVAYVTGV-QGGQIQVEANYAGNOSIGNYRCWFNPGSVSYIY 429

Db 212 QTSQGAYGHVAYVEGVNSNGSIRVSEMYGHGAGVVTSTRTISASQAASNY 262  
Search completed: February 10, 2006, 22:36:18  
Job time : 78.7712 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:14:04 ; Search time 252.865 Seconds  
(without alignments)  
712.176 Million cell updates/sec

Title: US-10-797-821-30  
Perfect score: 2104  
Sequence: 1 MKKRILSAVLVSGVTLLSAT.....SIGNRGWENPGSVSIYPN 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA\_Main:\*  
1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pgp:\*  
2: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pgp:\*  
3: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pgp:\*  
4: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pgp:\*  
5: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pgp:\*  
6: /cgn2\_6/prodata/1/pubpaa/US11\_PUBCOMB.pgp:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2104	100.0	431	US-10-383-930-30	Sequence 30, Appl
2	2104	100.0	431	US-10-797-821-30	Sequence 30, Appl
3	2100	99.8	431	US-10-383-930-29	Sequence 29, Appl
4	2100	99.8	431	US-10-797-821-29	Sequence 29, Appl
5	2096	99.6	431	US-10-383-930-33	Sequence 33, Appl
6	2096	99.6	431	US-10-797-821-33	Sequence 33, Appl
7	2083.5	99.0	432	US-10-383-930-31	Sequence 31, Appl
8	2083.5	99.0	432	US-10-797-821-31	Sequence 31, Appl
9	2074.5	98.6	432	US-10-383-930-32	Sequence 32, Appl
10	2074.5	98.6	432	US-10-797-821-32	Sequence 32, Appl
11	1085.5	51.6	398	US-10-474-792-600	Sequence 600, App
12	1057.5	50.3	392	US-10-472-928-4652	Sequence 4652, Ap
13	1057.5	50.3	399	US-10-617-320-3230	Sequence 3230, Ap
14	594	28.2	524	US-10-282-122A-57658	Sequence 57658, A
15	562	26.7	210	US-10-154-251-67	Sequence 67, Appl
16	503	23.9	449	US-09-071-035-482	Sequence 482, App
17	503	23.9	449	US-10-206-576-482	Sequence 482, App
18	503	23.9	449	US-10-912-362-482	Sequence 482, App
19	454	21.6	422	US-09-071-035-484	Sequence 484, App
20	454	21.6	422	US-10-206-576-484	Sequence 484, App
21	454	21.6	422	US-10-912-362-484	Sequence 484, App
22	337.5	16.0	630	US-10-494-674-6	Sequence 6, Appl1
23	331.5	15.8	600	US-09-738-626-5197	Sequence 5197, Ap
24	299	14.2	440	US-10-501-282-3056	Sequence 3056, Ap
25	299	14.2	609	US-10-501-282-3184	Sequence 3184, Ap
26	296.5	14.1	422	US-10-501-282-3054	Sequence 3054, Ap
27	287.5	13.7	377	US-10-501-282-3052	Sequence 3052, Ap

ALIGNMENTS

RESULT 1

US-10-383-930-30  
; Sequence 30, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 30  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
; US-10-383-930-30

Query Match	100.0%	Score	2104;	DB	4;	Length	431;
Best Local Similarity	100.0%	Pred. No.	3	6e-120;			
Matches	431;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	1	MKKRILSAVLVSGVTLLSAT	TTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG	60			
Db	1	MKKRILSAVLVSGVTLLSAT	TTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG	60			
Qy	61	QVSALOTQAELOAENQRL	EAQSATLGGQITQTLSKIVARNESLKQQAARSNAATSY	120			
Db	61	QVSALOTQAELOAENQRL	EAQSATLGGQITQTLSKIVARNESLKQQAARSNAATSY	120			
Qy	121	INAINSKVSDAINRVSAIR	EVVSVANEKMLQOEQDKAAVEQKQENQAINTVAANQE	180			
Db	121	INAINSKVSDAINRVSAIR	EVVSVANEKMLQOEQDKAAVEQKQENQAINTVAANQE	180			
Qy	181	TTAQTNTALNTQQAQL	EAQLNLOAELTTAQDQKATLVAQKAAAEAAARQAAAAA	240			
Db	181	TTAQTNTALNTQQAQL	EAQLNLOAELTTAQDQKATLVAQKAAAEAAARQAAAAA	240			
Qy	241	KAAAEKALQEQAAQAA	QAANNNTQATDASDQAAAAADNTQAAQTGDTSTEQSAAQVYNS	300			
Db	241	KAAAEKALQEQAAQAA	QAANNNTQATDASDQAAAAADNTQAAQTGDTSTEQSAAQVYNS	300			
Qy	301	QDSESTTATEAOP	SASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWAA	360			
Db	301	QDSESTTATEAOP	SASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWAA	360			

QY 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYVTGVGGQIQVOEANYAGNQSIGNYRGWF 420  
Db 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYVTGVGGQIQVOEANYAGNQSIGNYRGWF 420

QY 421 NPGSVSYIYPN 431  
Db 421 NPGSVSYIYPN 431

RESULT 2  
US-10-797-821-30  
; Sequence 30, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 30  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-30

Query Match 100.0%; Score 2104; DB 5; Length 431;  
Best Local Similarity 100.0%; Pred. No. 3.6e-120;  
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTACQQAQAQVNTIQG 60  
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTACQQAQAQVNTIQG 60

QY 61 QVSALOTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
Db 61 QVSALOTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQENQAIAINTVAANQE 180  
Db 121 INAIINSKVSDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQENQAIAINTVAANQE 180

QY 181 TTAQNTNALNTQQAQLLEAQLNLQAEELTTAQQDKATLVAQKAAAEAAARQAAAAA 240  
Db 181 TTAQNTNALNTQQAQLLEAQLNLQAEELTTAQQDKATLVAQKAAAEAAARQAAAAA 240

QY 241 KAAAEAKALQEQAQAQAQAANNNTQATDASDQQAADNTQAAQTGDSQSAQAQVNNNS 300  
Db 241 KAAAEAKALQEQAQAQAQAANNNTQATDASDQQAADNTQAAQTGDSQSAQAQVNNNS 300

QY 301 DOESTTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360  
Db 301 DOESTTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360

QY 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYVTGVGGQIQVOEANYAGNQSIGNYRGWF 420  
Db 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYVTGVGGQIQVOEANYAGNQSIGNYRGWF 420

QY 421 NPGSVSYIYPN 431  
Db 421 NPGSVSYIYPN 431

RESULT 3  
US-10-383-930-29  
; Sequence 29, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 29  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-29

Query Match 99.8%; Score 2100; DB 4; Length 431;  
Best Local Similarity 99.8%; Pred. No. 6.3e-120;  
Matches 430; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTACQQAQAQVNTIQG 60  
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTACQQAQAQVNTIQG 60

QY 61 QVSALOTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120  
Db 61 QVSALOTQQAELQAEQNRLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQENQAIAINTVAANQE 180  
Db 121 INAIINSKVSDAINRVSAIREVVSANERKMLQQEQDKAAVEQKQENQAIAINTVAANQE 180

QY 181 TTAQNTNALNTQQAQLLEAQLNLQAEELTTAQQDKATLVAQKAAAEAAARQAAAAA 240  
Db 181 TTAQNTNALNTQQAQLLEAQLNLQAEELTTAQQDKATLVAQKAAAEAAARQAAAAA 240

QY 241 KAAAEAKALQEQAQAQAQAANNNTQATDASDQQAADNTQAAQTGDSQSAQAQVNNNS 300  
Db 241 KAAAEAKALQEQAQAQAQAANNNTQATDASDQQAADNTQAAQTGDSQSAQAQVNNNS 300

QY 301 DOESTTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360  
Db 301 DOESTTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360

QY 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYVTGVGGQIQVOEANYAGNQSIGNYRGWF 420  
Db 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYVTGVGGQIQVOEANYAGNQSIGNYRGWF 420

QY 421 NPGSVSYIYPN 431  
Db 421 NPGSVSYIYPN 431

RESULT 4  
US-10-797-821-29  
; Sequence 29, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens

```
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PR
; ORGANISM: Streptococcus mutans
; ORGANISM: Streptococcus mutans
US-10-797-821-29

Query Match          99.8%; Score 2100; DB 5; Length 431;
Best Local Similarity 99.8%; Pred. No. 6.3e-120;
Matches 430; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSSTATLSAVKADDDFAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSSTATLSAVKADDDFAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

Qy 61 QVSALQTQAAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNAATSY 120
Db 61 QVSALQTQAAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNAATSY 120

Qy 121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQEODKAAVEQKQENQAINTVAANO 180
Db 121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQEODKAAVEQKQENQAINTVAANO 180

Qy 181 TIAQNTNALNTQQAELQAEALNLQAEALTAAQDQKATLVAQKAAAEAAEAARQAAAAEA 240
Db 181 TIAQNTNALNTQQAELQAEALNLQAEALTAAQDQKATLVAQKAAAEAAEAARQAAAAEA 240

Qy 241 KAAAEAKALQEAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNS 300
Db 241 KAAAEAKALQEAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNS 300

Qy 301 DQESTTATEAQPSSASTAAVAANTSSANTYPAGQCTGWKSLAPWVGNVWNGGOWAA 360
Db 301 DQESTTATEAQPSSASTAAVAANTSSANTYPAGQCTGWKSLAPWVGNVWNGGOWAA 360

Qy 361 SAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVGGQIQVQEAANYAGNQSIGNYRGWF 420
Db 361 SAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVGGQIQVQEAANYAGNQSIGNYRGWF 420

Qy 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431

RESULT 5
US-10-383-930-33
; Sequence 33, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08

; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PR
; ORGANISM: Streptococcus mutans
; ORGANISM: Streptococcus mutans
US-10-383-930-33

Query Match          99.6%; Score 2096; DB 4; Length 431;
Best Local Similarity 99.5%; Pred. No. 1.1e-119;
Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSSTATLSAVKADDDFAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSSTATLSAVKADDDFAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

Qy 61 QVSALQTQAAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNAATSY 120
Db 61 QVSALQTQAAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNAATSY 120

Qy 121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQEODKAAVEQKQENQAINTVAANO 180
Db 121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQEODKAAVEQKQENQAINTVAANO 180

Qy 181 TIAQNTNALNTQQAELQAEALNLQAEALTAAQDQKATLVAQKAAAEAAEAARQAAAAEA 240
Db 181 TIAQNTNALNTQQAELQAEALNLQAEALTAAQDQKATLVAQKAAAEAAEAARQAAAAEA 240

Qy 241 KAAAEAKALQEAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNS 300
Db 241 KAAAEAKALQEAQAQAANNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNNS 300

Qy 301 DQESTTATEAQPSSASTAAVAANTSSANTYPAGQCTGWKSLAPWVGNVWNGGOWAA 360
Db 301 DQESTTATEAQPSSASTAAVAANTSSANTYPAGQCTGWKSLAPWVGNVWNGGOWAA 360

Qy 361 SAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVGGQIQVQEAANYAGNQSIGNYRGWF 420
Db 361 SAAAAGYRVGSTPSAGAVAVMNDGGYGHVAVYTVGVGGQIQVQEAANYAGNQSIGNYRGWF 420

Qy 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431

RESULT 6
US-10-797-821-33
; Sequence 33, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
```

```
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-33

Query Match      99.6%; Score 2096; DB 5; Length 431;
Best Local Similarity 99.5%; Pred. No. 1.1e-119;
Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFPAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFPAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
Db 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120

Qy 121 INAIINSKVSVDAINRVSAIREVVSANEKMLQQQEQDKAAVEQKQENQAINTVAANQE 180
Db 121 INAIINSKVSVDAINRVSAIREVVSANEKMLQQQEQDKAAVEQKQENQAINTVAANQE 180

Qy 181 TIAQNTNALNTQQAELQAEOLNLOAELTTAQQOKATLVAQKAAAEBAARQAAQAQAAEA 240
Db 181 TIAQNTNALNTQQAELQAEOLNLOAELTTAQQOKATLVAQKAAAEBAARQAAQAQAAEA 240

Qy 241 KAAAEAKALQEQAAQAAQAAA - NNNTQATDASDQAAADNTQAAQTGDSSTEQSAAQAVNN 299
Db 241 KAAAEAKALQEQAAQAAQAAA - NNNTQATDASDQAAADNTQAAQTGDSSTEQSAAQAVNN 299

Qy 300 SDESTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWAA 359
Db 300 SDESTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGQWAA 359

Qy 360 ASAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYVTGVGGQIQVQEAANYAGNQSIGNYRGW 419
Db 360 ASAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYVTGVGGQIQVQEAANYAGNQSIGNYRGW 419

Qy 420 FNPGSVSYIYPN 431
Db 420 FNPGSVSYIYPN 432

RESULT 8
US-10-797-821-31
; Sequence 31, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-31

Query Match      99.0%; Score 2083.5; DB 5; Length 432;
Best Local Similarity 99.1%; Pred. No. 6.4e-119;
Matches 428; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Qy 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFPAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFPAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
Db 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120

Qy 121 INAIINSKVSVDAINRVSAIREVVSANEKMLQQQEQDKAAVEQKQENQAINTVAANQE 180
Db 121 INAIINSKVSVDAINRVSAIREVVSANEKMLQQQEQDKAAVEQKQENQAINTVAANQE 180
```



```
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOEQDKAAVEQKQENQAAINTVAANQE 180
QY 181 TIAQNTNALNTQQAQLEAAQNLQAEITTAQDQKATILVAQKAAAEAAARQAAAAQAAAAE 240
Db 181 TIAQNTNALNTQQAQLEAAQNLQAEITTAQDQKATILVAQKAAAEAAARQAAAAQAAAAE 240
QY 241 KAAAEAKALQEQAAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299
Db 241 KAAAEAKALQEQAAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300
QY 300 SDOESTTATEAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWVGYNGGQWA 359
Db 301 SDOESTTATEAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWVGYNGGQWA 360
QY 360 ASAAAAGYRVGSTPSAGAVVNDGGYGHVAVYTVGGGQIQVEANYAGNQSIGNYRGW 419
Db 361 ASAAAAGYRVGSTPSAGAVVNDGGYGHVAVYTVGGGQIQVEANYAGNQSIGNYRGW 420
QY 420 FNPGSVSYIYPN 431
Db 421 FNPGSVSYIYPN 432

RESULT 9
US-10-383-930-32
; Sequence 32, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-32

Query Match 98.6%; Score 2074.5; DB 4; Length 432;
Best Local Similarity 98.8%; Pred. No. 2.2e-118;
Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQ 60
QY 61 QVSALOTQQAELQAEQNRLEAQSAATLGQOIQTLLSSKIVARNESLKQQAARSQAQSNAAATSY 120
Db 61 QVSALOTQQAELQAEQNRLEAQSAATLGQOIQTLLSSKIVARNESLKQQAARSQAQSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOEQDKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOEQDKAAVEQKQENQAAINTVAANQE 180
QY 181 TIAQNTNALNTQQAQLEAAQNLQAEITTAQDQKATILVAQKAAAEAAARQAAAAQAAAAE 240
Db 181 TIAQNTNALNTQQAQLEAAQNLQAEITTAQDQKATILVAQKAAAEAAARQAAAAQAAAAE 240
QY 241 KAAAEAKALQEQAAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299
Db 241 KAAAEAKALQEQAAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300
QY 300 SDOESTTATEAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWVGYNGGQWA 359
Db 300 SDOESTTATEAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWVGYNGGQWA 359
QY 360 ASAAAAGYRVGSTPSAGAVVNDGGYGHVAVYTVGGGQIQVEANYAGNQSIGNYRGW 419
Db 360 ASAAAAGYRVGSTPSAGAVVNDGGYGHVAVYTVGGGQIQVEANYAGNQSIGNYRGW 420
```

```
Db 301 SDOESTTATEAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWVGYNGGQWA 360
QY 360 ASAAAAGYRVGSTPSAGAVVNDGGYGHVAVYTVGGGQIQVEANYAGNQSIGNYRGW 419
Db 361 ASAAAAGYRVGSTPSAGAVVNDGGYGHVAVYTVGGGQIQVEANYAGNQSIGNYRGW 420
QY 420 FNPGSVSYIYPN 431
Db 421 FNPGSVSYIYPN 432

RESULT 10
US-10-797-821-32
; Sequence 32, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-32

Query Match 98.6%; Score 2074.5; DB 5; Length 432;
Best Local Similarity 98.8%; Pred. No. 2.2e-118;
Matches 427; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAQAAQVNTIQ 60
QY 61 QVSALOTQQAELQAEQNRLEAQSAATLGQOIQTLLSSKIVARNESLKQQAARSQAQSNAAATSY 120
Db 61 QVSALOTQQAELQAEQNRLEAQSAATLGQOIQTLLSSKIVARNESLKQQAARSQAQSNAAATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOEQDKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQOEQDKAAVEQKQENQAAINTVAANQE 180
QY 181 TIAQNTNALNTQQAQLEAAQNLQAEITTAQDQKATILVAQKAAAEAAARQAAAAQAAAAE 240
Db 181 TIAQNTNALNTQQAQLEAAQNLQAEITTAQDQKATILVAQKAAAEAAARQAAAAQAAAAE 240
QY 241 KAAAEAKALQEQAAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299
Db 241 KAAAEAKALQEQAAQAAAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300
QY 300 SDOESTTATEAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWVGYNGGQWA 359
Db 301 SDOESTTATEAOPSASSASTAAVAANTSSANTYPAGCTGWGKSLAPWVGYNGGQWA 360
QY 360 ASAAAAGYRVGSTPSAGAVVNDGGYGHVAVYTVGGGQIQVEANYAGNQSIGNYRGW 419
Db 361 ASAAAAGYRVGSTPSAGAVVNDGGYGHVAVYTVGGGQIQVEANYAGNQSIGNYRGW 420
```

```
QY 420 FNPQSVSYIYPN 431
Db 421 FNPQSVSYIYPN 432
Db 421 FNPQSVSYIYPN 432

RESULT 11
US-10-474-792-600
; Sequence 600, Application US/10474792
; Publication No. US20040236072A1
; GENERAL INFORMATION:
; APPLICANT: Olmsted, Stephen
; APPLICANT: Zagursky, Robert
; APPLICANT: Nickbarg, Elliot
; APPLICANT: Winter, Lourie
; TITLE OF INVENTION: SURFACE PROTEINS OF STREPTOCOCCUS PYOGENES
; FILE REFERENCE: AM 100399
; CURRENT APPLICATION NUMBER: US/10/474,792
; CURRENT FILING DATE: 2003-10-14
; NUMBER OF SEQ ID NOS: 674
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 600
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-10-474-792-600

Query Match 51.6%; Score 1085.5; DB 5; Length 398;
Best Local Similarity 53.0%; Pred. No. 3.2e-58;
Matches 229; Conservative 60; Mismatches 108; Indels 35; Gaps 5;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNATSY 120
Db 58 QVSSLOEQDKLTARNTLEALSFRPEQBIKALTSQIVARNEKLNKQQAARSAYKNNETSGY 117
QY 121 INAINSKSVDAINRVSAIREVVSANERKMLQOQODKAAVEKQOENQAQVNTVAANQE 180
Db 118 INALLSKSISDVNVLAINRAVSAVNAKLLLEQKADKVSLEEKQAANQTAINNTIAANWA 177
QY 181 TTAQNTNALNTQQAOLAEALQALNLAQELTTAODQKATLVAKAAAEAEAAQAAAAAEEA 240
Db 178 MAEENQNTLTQQAANLVAATANLALQASATEDKANLVQKEAAEKAAAEALAEQAQAKV 237
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSTEQSAQAQVNNNS 300
Db 238 KAAEQEA---AQQAASVEAKSAITPAQ-----TPAAQSSNAIEPAAL----- 278
QY 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGOWAA 360
Db 279 -----TAPAAPSAGPQTSYDSSNTYPVQGCTWGAKS LAPWAGNNWNGGOWAY 326
QY 361 SAAAGYRGSTPSAGAVAVWNDGGYGHVAVYTVGO-GGOIOVOEANYAGNOSIGNYRGW 419
Db 327 SAQAAGYRGSTPSAGAVAVWNDGGYGHVAVVVEVQSASSIRWESNYSGRQYIADHRGW 386
QY 420 FNPQSVSYIYPN 431
Db 387 FNPQSVSYIYPN 398

RESULT 12
US-10-472-928-4652
; Sequence 4652, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926W0
```

```
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 4652
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: secreted 45 kd protein (usp45)
; OTHER INFORMATION: Cellular location: outside
; OTHER INFORMATION: Similar to strain R6 sequence 15904062 (O.E+01)
US-10-472-928-4652

Query Match 50.3%; Score 1057.5; DB 5; Length 392;
Best Local Similarity 50.3%; Pred. No. 1.6e-56;
Matches 219; Conservative 73; Mismatches 92; Indels 51; Gaps 5;

QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
Db 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIOG 60
QY 61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQKSNATSY 120
Db 61 QVSAIQAEOSNLQAEENDRLQAESKKLEGITELSKNIVSRNQSLERKQARSQAQVNTIOG 120
QY 121 INAINSKSVDAINRVSAIREVVSANERKMLQOQODKAAVEKQOENQAQVNTVAANQE 180
Db 121 INTIVNSKITEAISRAVVAAMSEIVSANNKMLEQKQADKKAISEKQVANNDAINTVIANQQ 180
QY 181 TTAQNTNALNTQQAOLAEALQALNLAQELTTAODQKATLVAKAAAEAEAAQAAAAAEEA 240
Db 181 KLADDAQAALTQQAELKAAELSLAAEKATAEGEKASLLEQKAAAEAEAAVAAEAYK- 239
QY 241 KAAAEAKALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSTEQSAQAQVNNNS 300
Db 240 -----EKRAQQQSVLASANTNLTAQVQAVSESAAA----- 270
QY 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGOWAA 360
Db 271 ----PVRAKVRPTYS-----TNASSYPIGECTGWGKTLAPWAGDYGNGAQWAT 315
QY 361 SAAAGYRGSTPSAGAVAVWNDGGYGHVAVYTVGO-GGOIOVOEANYAGNOSIGNYRGW 419
Db 316 SAAAGYRGSTPSAGAVAVWNDGGYGHVAVVAVESTTRIQVSESNYAGNRTIGHRGW 375
QY 420 FNP-----GSVSYIY 429
Db 376 FNPPTTSEGFTVIY 390

RESULT 13
US-10-617-320-3230
; Sequence 3230, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
```

```
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/10/617,320
;   FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US/09/107,433
;   FILING DATE: 30-Jun-1998
;   APPLICATION NUMBER: 60/ 085131
;   FILING DATE: May 12, 1998
;   APPLICATION NUMBER: 60/051553
;   FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
;   NAME: Ariniello, Pamela Deneke
;   REGISTRATION NUMBER: 40,489
;   REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (781)893-5007
;   TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3230:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 399 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
;     HYPOTHETICAL: YES
;     ORIGINAL SOURCE:
;       ORGANISM: Streptococcus pneumoniae
;     FEATURE:
;       NAME/KEY: misc feature
;       LOCATION: (B) LOCATION 1...399
;     SEQUENCE DESCRIPTION: SEQ ID NO: 3230:
US-10-617-320-3230

Query Match          50.3%; Score 1057.5; DB 5; Length 399;
Best Local Similarity 50.3%; Pred. No. 1.6e-56;
Matches 219; Conservative 73; Mismatches 92; Indels 51; Gaps 5;

Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Db 8 MKKILASLLSTVMVSGVAVLTTHAETDDKIAAQDNKISNLTAAQQEAQKQVDQIQE 67
Qy 61 QVSALQTQQAELQENORLEAQSATLGGQITLSSKIVARNESLKQQAARSQAQSNATSY 120
Db 68 QVSAIQAEQSNLQENRLEAQSATLGGQITLSSKIVARNESLKQQAARSQAQSNATSY 127
Qy 121 INATINSKSVSDAINRVSAIREVVSAANEKMLQQEQDQKAAVEQKQQAARSQAQSNATSY 180
Db 128 INTIVNSKSTEARLSRVAAVSEIVSANNKLEQQQKAKKALSEKQVANNDAINTVIANQQ 187
Qy 181 TIAQNTNALNTQQAELAAQLNLQAEELTTAQDQKATLVAAQKAAAEAAQAAQAAAE 240
Db 188 KLADDAQALTTKQAEKAAELSLAAEKATABEGEKASLLEQKAAAEAAAEAAVAAVK- 246
Qy 241 KAAAEKALQEAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSQSAQAQVNN 300
Db 247 ----EKRAEQQSQVLAANTNLTAQVQAVSESA-- 277
Qy 301 DQESTTATEQPSASSASTAAVAANTSSANTYPAGQCTGWKSLAPVGNVWNGGOWAA 360
Db 278 ----PVRAKVRPTYS-----TNASSYPGECTGWKTLAPWAGDVTWNGAQWAT 322
Qy 361 SAAAGYRVGSTPSAGAVAVMNDGGYGHVAVTGVQGG-QGTQVQEAANYAGNOSTGNRGW 419
Db 323 SAAAGFRGTSTPQVGAICWNDGGYGHVAVTAVESTTRIQVESYVAGNRTIGNRGW 382
Qy 420 FNP-----GSVSYTY 429
Db 383 FNPPTTSEGFTYTY 397

RESULT 14
US-10-282-122A-57658
```

```
; Sequence 57658, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57658
; LENGTH: 524
; TYPE: PRT
; ORGANISM: Enterococcus faecium
; US-10-282-122A-57658
```

```
Query Match          28.2%; Score 594; DB 4; Length 524;
Best Local Similarity 31.0%; Pred. No. 3.7e-28;
Matches 162; Conservative 91; Mismatches 164; Indels 106; Gaps 14;

Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60
Db 1 MKKLSLSAVVWCMTLTAVASPTAAADDFDSQIQDQDKIADLKNOQADQASQIDALE 60
Qy 61 QVSALQTQQAELQENORLEAQSATLGGQITLSSKIVARNESLKQQAARSQAQSNATSY 120
Db 61 QVSEINTQADLLAKQDQLRQESAQLVKDIADLQERIEKREDTIQKQAREAQVNTSSNY 120
Qy 121 INATINSKSVSDAINRVSAIREVVSAANEKMLQQEQDQKAAVEQKQQAARSQAQSNATSY 180
Db 121 IDAVLNADSLADIAIGRVQAMTTVMKANNDLMEQKQDKKAVEDKKAENDAKLKEAENQA 180
Qy 181 TIAQNTNALNTQQAELAAQLNLQAEELTTAQDQKATLVAAQKAAAEAAEAAE 226
Db 181 ALESQKGDLLSKQADLNLVLTSLAAEQATAEKADLNROKAAEAQAEQAEQAEQAEQAEQ 240
Qy 227 AARQAAAAQAAAEAKAAAEAAKALQEQAAQ-AQAAANNNTQATDAS-----DQAAAAADNTQ 281
Db 241 ARQQAQAEKAEKAEQAEAAEQATQASSTAQSSATESSATQSSMTTESSSATQSSATE 300
Qy 282 AATGDSQSAQAQVNNNSDQESTTATEA-----QPSASSASTAAVAANTSSANT 331
```

```
Db      301 ESTTPSTESTAPESATEESTTAPESATEESTTVPESATEESTTVPESSTTEEST 360
      332 YPA-----GCTWGVKSLAP----- 346
Db      361 TPAPTTPTDQSVDTGNGT-GSSTPAPTPTTPEQPKVTPAPAPSGSVNGAAIVAEAYK 419
      347 WVGNG--YWGNG-----GOWAASAAAAGYRVG-STPSAGAVAV 380
Db      420 YIGTPYWGKDPGPGPCSFTRVYVMQVTRGRDIGTVPQESAGTKISVSQAKAGDLLF 479
      381 W-NDGGYGHVAYVTGQGGIQVQOEANYAGNQSIGNYRGWFPN 422
Db      480 WSGQGGYTHVAIALG--GGQ-YIHAPQPGESVKGVSVQ-WFAP 518
```

```
RESULT 15
US-10-154-251-67
; Sequence 67, Application US/10154251
; Publication No. US20030092024A1
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Frittz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 06286-060002
; CURRENT APPLICATION NUMBER: US/10/154,251
; PRIOR FILING DATE: 2002-09-16
; PRIOR APPLICATION NUMBER: US 10/154,251
; PRIOR FILING DATE: 2002-05-22
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 67
; LENGTH: 210
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-154-251-67
```

```
Query Match      26.7%; Score 562; DB 4; Length 210;
Best Local Similarity 55.2%; Pred. No. 1.2e-26;
Matches 116; Conservative 45; Mismatches 49; Indels 0; Gaps 0;

Qy      1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAQIASODSKINNLTAAQQAAQAQVNTIQG 60
Db      1 MKKRILASLSTVWVSQVAVLTTHAETTTDDKIAQDNKISNLTAAQQQAQKQVQIQE 60

Qy      61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKKQARSQAQKSNATSY 120
Db      61 QVSAIQAEQSNLQAEENDRLQAESKKLEGEITELSKNIVSRNQSLKQARSQAQTNGAVTSY 120

Qy      121 INAIINSKVSDAINRVSAIRVVSVANEKMLQQEQDQKAAVEQKQENQAAINTVAANOE 180
Db      121 INTIVNSKSITEAIRVAAMSEIVSANNKMLEQKADKKAISEKQVANNDAINVTIANQQ 180

Qy      181 TIAQNTNALNTQQAQLEAAQINLQAEITTA 210
Db      181 KLADDAQALTTKQAEKLAELSLAAEKATS 210
```

Search completed: February 10, 2006, 23:25:15  
Job time : 253.865 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:15:51 ; Search time 18.6825 Seconds  
(without alignments)  
302.734 Million cell updates/sec

Title: US-10-797-821-30  
Perfect score: 2104  
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNRGWFGNSVSIYPN 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New:  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2100	99.8	431	7	US-11-052-554A-210
2	1092.5	51.9	398	7	US-11-052-554A-252
3	435	20.7	211	7	US-11-052-554A-352
4	410.5	19.5	544	7	US-11-052-554A-358
5	233	11.1	257	6	US-10-793-626-3244
6	230	10.9	394	7	US-11-052-554A-79
7	227.5	10.8	971	7	US-11-052-554A-7
8	224.5	10.7	1236	6	US-10-873-528-109
9	215	10.2	149	6	US-10-793-626-1682
10	215	10.2	157	6	US-10-793-626-2870
11	213	10.1	270	6	US-10-485-517-413
12	204	9.7	1562	7	US-11-052-554A-211
13	195	9.3	1095	6	US-10-793-626-3154
14	194.5	9.2	2101	6	US-10-857-780-23
15	192.5	9.1	1448	6	US-10-485-517-212
16	190	9.0	1586	6	US-10-821-234-901
17	188	8.9	1410	6	US-10-878-556A-136
18	186	8.8	1107	6	US-10-485-517-145
19	185.5	8.8	758	6	US-10-485-517-144
20	185	8.8	1122	6	US-10-467-657-6112
21	180.5	8.6	1290	6	US-10-485-517-141
22	179.5	8.5	330	6	US-10-485-517-415
23	177	8.4	3712	7	US-11-019-711-48
24	177	8.4	3712	7	US-11-019-711-51
25	175.5	8.3	1126	6	US-10-485-517-248

RESULT 1  
US-11-052-554A-210  
; Sequence 210, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 210  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans UA159  
US-11-052-554A-210

ALIGNMENTS

26	173.5	8.2	834	7	US-11-052-554A-212	Sequence 212, App
27	173	8.2	1565	6	US-10-467-657-2704	Sequence 2704, App
28	173	8.2	5024	6	US-10-793-626-2964	Sequence 2964, App
29	171.5	8.2	1992	7	US-11-069-834-58	Sequence 58, Appl
30	171.5	8.2	2000	7	US-11-069-834-56	Sequence 56, Appl
31	166.5	7.9	761	6	US-10-485-517-252	Sequence 252, App
32	166.5	7.9	1634	7	US-11-019-711-49	Sequence 49, Appl
33	166.5	7.9	1640	7	US-11-019-711-8	Sequence 8, Appli
34	166.5	7.9	3690	6	US-10-995-561-1016	Sequence 1016, Ap
35	166.5	7.9	3714	6	US-10-995-561-1015	Sequence 1015, Ap
36	166.5	7.9	3717	6	US-10-821-234-1076	Sequence 1076, Ap
37	165	7.8	860	7	US-11-019-711-59	Sequence 59, Appl
38	164.5	7.8	612	6	US-10-467-657-3988	Sequence 3988, Ap
39	164.5	7.8	706	6	US-10-485-517-146	Sequence 146, App
40	162	7.7	684	6	US-10-793-626-2098	Sequence 2098, Ap
41	158.5	7.5	365	7	US-11-052-554A-223	Sequence 223, App
42	157.5	7.5	716	7	US-11-194-890-20	Sequence 20, Appl
43	156.5	7.4	1404	6	US-10-878-556A-169	Sequence 169, App
44	156.5	7.4	1995	7	US-11-069-834-60	Sequence 60, Appl
45	156	7.4	655	6	US-10-793-626-1052	Sequence 1052, Ap

Query Match		99.8%	Score	2100;	DB 7;	Length	431;
Best Local Similarity		99.8%	Pred. No.	7.8e-113;			
Matches		430;	Conservative	0;	Mismatches	1;	Indels 0; Gaps 0;
Qy	1	MKKRILSAVLVSGVTLSAT	1	MKKRILSAVLVSGVTLSAT	1	MKKRILSAVLVSGVTLSAT	1
Db	1	MKKRILSAVLVSGVTLSAT	1	MKKRILSAVLVSGVTLSAT	1	MKKRILSAVLVSGVTLSAT	1
Qy	61	QVSALQTQQAELQAEORLEAQSATL	61	QVSALQTQQAELQAEORLEAQSATL	61	QVSALQTQQAELQAEORLEAQSATL	61
Db	61	QVSALQTQQAELQAEORLEAQSATL	61	QVSALQTQQAELQAEORLEAQSATL	61	QVSALQTQQAELQAEORLEAQSATL	61
Qy	121	INAIINSKVSVDAINRVSAIRVVSA	121	INAIINSKVSVDAINRVSAIRVVSA	121	INAIINSKVSVDAINRVSAIRVVSA	121
Db	121	INAIINSKVSVDAINRVSAIRVVSA	121	INAIINSKVSVDAINRVSAIRVVSA	121	INAIINSKVSVDAINRVSAIRVVSA	121
Qy	181	TIANTNTALNTQQAELQAEORLEA	181	TIANTNTALNTQQAELQAEORLEA	181	TIANTNTALNTQQAELQAEORLEA	181
Db	181	TIANTNTALNTQQAELQAEORLEA	181	TIANTNTALNTQQAELQAEORLEA	181	TIANTNTALNTQQAELQAEORLEA	181
Qy	241	KAAAEAKALQEAQAAQAAANNTQ	241	KAAAEAKALQEAQAAQAAANNTQ	241	KAAAEAKALQEAQAAQAAANNTQ	241
Db	241	KAAAEAKALQEAQAAQAAANNTQ	241	KAAAEAKALQEAQAAQAAANNTQ	241	KAAAEAKALQEAQAAQAAANNTQ	241
Qy	301	DQESTTTEAQPSASSASTAAVAANT	301	DQESTTTEAQPSASSASTAAVAANT	301	DQESTTTEAQPSASSASTAAVAANT	301
Db	301	DQESTTTEAQPSASSASTAAVAANT	301	DQESTTTEAQPSASSASTAAVAANT	301	DQESTTTEAQPSASSASTAAVAANT	301

```
Db 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVGNVWNGGGQWAA 360
Qy 361 SAAAAGRVGSTPSAGAVAVWNDGGYGHVAVTVTGQGGQIQVQAEYANTAGNQSIGNYRGWF 420
Db 361 SAAAAGRVGSTPSAGAVAVWNDGGYGHVAVTVTGQGGQIQVQAEYANTAGNQSIGNYRGWF 420
Qy 421 NPGSVSVIYPN 431
Db 421 NPGSVSVIYPN 431

RESULT 2
US-11-052-554A-252
; Sequence 252, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 252
; LENGTH: 398
; TYPE: PR1
; ORGANISM: Streptococcus pyogenes MGAS8232
US-11-052-554A-252

Query Match 51.9%; Score 1092.5; DB 7; Length 398;
Best Local Similarity 53.7%; Pred. No. 1.2e-55;
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;

Qy 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 60
Db 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAVNTIQG 57

Qy 61 QVSLQGEQDKLTARTELEALSIRFEQIKALTSQIVARNEKLNQARSAYKNNETSGY 117
Db 58 QVSLQGEQDKLTARTELEALSIRFEQIKALTSQIVARNEKLNQARSAYKNNETSGY 117

Qy 121 INALINSKVSDAINRVSAIREVVSANEKMLQOQEDKAAVEQKQENQAAINTVAANQE 180
Db 118 INALLNSKSIDVNVRLVAINRAVSANAKLLEQKADKVSLEEKQAAANTAIANMA 177

Qy 181 TIAQNTWALNTQQQLBAQLNLQELTTQDDQKATLVAKOAAAEAAARQAAQAAAE 240
Db 178 MABENQNTLTQQANLSAANTANLALQASATEDKANLVAKQEAEEKAAAEALAQEAAKV 237

Qy 241 KAAAEAKALQEQRAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSSTEQSAQVNS 300
Db 238 KAQEQA---ACQAAASVEAAKSATTPAQ-----TPAAQSSNAIEPAALTA---- 280

Qy 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVGNVWNGGGQWAA 360
Db 281 ----PAAPSAARPQTS-----YDSNTYPVPGQCTGWGKSLAPVGNVWNGGGQW 326

Qy 361 SAAAAGRVGSTPSAGAVAVWNDGGYGHVAVTVTGQGGQIQVQAEYANTAGNQSIGNYRGW 419
Db 327 SAQAAGYRTGSTPMVGAIAVWNDGGYGHVAVTVTGQGGQIQVQAEYANTAGNQSIGNYRGW 386

Qy 420 NPGSVSVIYPN 431
Db 387 FNPTGVTFIYPH 398

RESULT 3
```

```
US-11-052-554A-352
; Sequence 352, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 352
; LENGTH: 211
; TYPE: PR1
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-352

Query Match 20.7%; Score 435; DB 7; Length 211;
Best Local Similarity 44.4%; Pred. No. 1.2e-18;
Matches 95; Conservative 28; Mismatches 57; Indels 34; Gaps 7;

Qy 247 KALQEQAAQAAQAAANNNTQATDASDQQAADNTQAAQTGDSSTEQSAQ-----AV 297
Db 3 KQFLEKAVFTVAA-----TAATVVLGNKMDAD-TYTLQEGDSFFSVQAQRVHMDAYELASM 57

Qy 298 NNSDQES-----TTATEAQPSSASSAST-AAVAANTSSANTYPAGQCTGWGKSL 344
Db 58 NGKDI TSLILPGQTLTVNGSAAPDNQAAAPTDTTQATTENDANANTYPVGCCTGWGKAV 117

Qy 345 APWVGNVWNGGQWAAASAAAGYRVGSTPSAGAVAVWNDGGYGHVAVTVGV-QGGQIQVQ 403
Db 118 ATWAGDWNGGWDWASSASAGQYTVGNTPAVGSMCTWDGGYGHVAVTVAVGDKGVQL 177

Qy 404 EANYAGNQSIGNYRGWPN-----PGSVSVIYPN 431
Db 178 ESNYKQDQWVDNRYRGWFDPNNSGTPGVSIVYPN 211

RESULT 4
US-11-052-554A-358
; Sequence 358, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 358
; LENGTH: 544
; TYPE: PR1
; ORGANISM: Streptococcus mutans UA159
US-11-052-554A-358

Query Match 19.5%; Score 410.5; DB 7; Length 544;
Best Local Similarity 25.3%; Pred. No. 8.9e-17;
Matches 142; Conservative 74; Mismatches 193; Indels 153; Gaps 17;

Qy 4 RILSAVLVSGVTLSATLTSVAVKADDFDAQIA-----SQDSKINNLTAAQQAAQAVNTIQ 59
Db 2 RKLKVALFASSILGMLAVSSYTAADTQNTSHYNEQAQTFDVAQAAQNGKTIQSID 61
```

```
QY 60 GOVSALQTOQAELQ-----AENORLEAQS-----ATL 86
Db 62 VAIWSEENGQDDLKWYHASNDSNQLTVHFNAENHSGSKVGSYIAHAYITYTDGNRVGNL 121
QY 87 GQOIOTLSKSVARNESLKQ---QARSQAQKNAATSYINAIINSKVSVDAINRVSA--- 139
Db 122 GKRKLKLSL---APOLSLKQGLQLFSLKPSAADQLFSVWSDENGQDDLHWYTDADG 177
QY 140 -----IREVVSANEKMLQQDEQD---KAAVE-QKQENQAAINTVAAN 178
Db 178 NTLAGYANHKGVTYVHVHTYILKONGKMPISAQDIDIPKPKVKIQIDKINDTSYDVVWN 237
QY 179 -----QETIAQNTNALNTQQAQLEAAQNLQAE-- 206
Db 238 VPPYISSVAIPWSEONGQDDLKWYQATKVADGIFKFTTVYKTRFELNGYQAHYIGDSQ 297
QY 207 -----LTTAQDQKATL---VAQKAAAEERARQAAAAQAAAEAKAAAEA 246
Db 298 LSKKLDGLGETHFNVPSIINVEDPQVTTIDHYNINKGTFDVTVAETDNSKAIQSISAAVWS 357
QY 247 KALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTG-----DSTEQSAQ 295
Db 358 DANQANLYWEAKQLANGKAAITVDVQ-----KHGNQIGSYNVHVYVYNDGTTSGHVL 411
QY 296 AVNNSDQESTTATEAOPSASSASTAAVAANTSSANTYPAGQCTGKVSGLAPWGVNNG 355
Db 412 A---NQLNQIVHYQPSA-----VRITAYVMEKNTPYVGQCTGKVKELAPWPNLNG 462
QY 356 GOWAASAAAAGYRVGSTPSAGAVAVWVNDGGYGHVAVYVTVGO-QGOIOVQEAANYAGNOSIG 414
Db 463 QOWASTVAVRGKIGTVPKVGAICWSDGGYGHVAVYVTVHVESNNRIQVKEANYKNOQYIS 522
QY 415 NYRGWFNP-----GSVSYIYPN 431
Db 523 NFRGWEDPTTSYLGRLTYIYD 544

RESULT 5
US-10-793-626-3244
; Sequence 3244, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
US-10-793-626-3244

Query Match 11.1%; Score 233; DB 6; Length 257;
Best Local Similarity 34.5%; Pred. No. 4.6e-07;
Matches 59; Conservative 29; Mismatches 75; Indels 8; Gaps 5;

261 NNNTQATDASQAAAAADNTQAAQTGSTESQASAAQAVNNSDOESTTATEAOPSASSASTA 320
91 NNYNNYNNYNNY---SNNTSQRTQPTQGLGASYSSTSSNVHVTTISA-PSNGVSL 146
321 AVAANTSANTYPAGQCTGWY-KSLAPWGVNNGGOWAASAAAAGYRVGSTPSAGAVA 379
147 --NARSAGNLYTSGQCTYVYVDFRGGKIGTWGNWANNWANAARSGYTVNNSPAKGAIL 204
```

```
QY 380 VVNDGGYGHVAVYTVGV-QGGOIOVQEAANYAGNOSIGNYRGWFPNPGSVSYIY 429
Db 205 QTSQAGYGHVAYVEGVNSNGSIRVSEMNYGHGAGVVTSTRTISASQAASNY 255

RESULT 6
US-11-052-554A-79
; Sequence 79, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent In version 3.3
; SEQ ID NO 79
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-79

Query Match 10.9%; Score 230; DB 7; Length 394;
Best Local Similarity 27.5%; Pred. No. 1.1e-06;
Matches 109; Conservative 63; Mismatches 176; Indels 48; Gaps 15;

5 ILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQ-----DSKINNLTAAQQAQAVNTIQ 59
16 IISAVL--HVILFAALIWS-----FDENIEASAGGGGSSIDAVMVDSGAVVEQYKRMQ 68
60 GOVSALQTOQAELQAEQNRLEAQSATLQCOQTQTLSSKIVARNESLKQ-----QARSQAQS 114
69 SQESS-----AKRSDEQRKMEQQA-----EELREKQAAAEQERLKQLEKRLAAEQEKK 118
115 NAATSYINAIINSKVSDAINRVSAIREVVSANEKMLQQOQODKAAVQKQOENQAAINT 174
119 QAEEAAKQAEELKQQAEEAAKAAADAKAKAEADDKAAEAAKAAADAKKAAEAAKA 178
175 VAANQETIAQNTNALNTQQAQLEAAQLNLQ---AELTTAQDQKATLVAKKAAAEBA--R 229
179 AAEQAQKAAEAAAALKKKAAEAAEAAAEAAEAKKAAEAAADKKA---AEKAAAEKAAADK 235
230 QAAAAQAAAAEAKAAEAKALQEQAAQAAQAAANNNTQATDASDQAAAAADNTQAAQTGST 289
236 KAAAEKAAADKKA---AKAAAEKAAAKAAA-----EADDIFGELSSGKNAPKTGGGA 287
290 EQSAAQAVNNSDOESTTATEAQ-PSASSASTAAVAANTSSANTYPAGQCTGWKVS LAP-- 346
288 KGNASPAGSGNTKNGASGADINNYAQIKSAIESKESFYDASSYAGTKTCTLRK-LAPDG 346
347 WVGNYWNGGOWA-ASAAAAGYRVGSTPSAGAVAVW 381
347 MLLDIKPEGDPALCQALAAALAAKLAIPKPPSQAVY 382

RESULT 7
US-11-052-554A-3
; Sequence 3, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
```

```
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
; LENGTH: 971
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-3

Query Match      10.8%; Score 227.5; DB 7; Length 971;
Best Local Similarity 25.5%; Pred. No. 4.2e-06;
Matches 92; Conservative 58; Mismatches 172; Indels 39; Gaps 8;

Qy 19 ATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTTQGVSAIQVLAQAEQENQR 78
Db 121 AONTAAAKKSADASTSAREAAATATDAADSARAASTSAGQAASQASASSAGTASTKA 180
Qy 79 LEAQSATLGGQIQITLSSKIVARNESLKQARSQAKSNAATSYINAIINSKSVDAINRVS 138
Db 181 TEA-----SKSAAAESSKGAATSGA-AKTSETNAVVSQGSAA-----TS 221
Qy 139 AIREVVSANERKMLQOQBDKAAVEQKQOENQAAINTVAANQETIAQNTNALNTQQAQLEA 198
Db 222 ASTATTKASEAASARDASASKEAAKSSSETSAASSASSA-----ASSATAAGNSAKAAKT 276
Qy 199 AQLNLQALHTTAQDKATLVAKAAAEARQAAAAQAAAEAKAAAEKALQEQAAQAA 258
Db 277 SETNAKSETAAEQSAAAGSKTAAALSAASASTSAGQASASATAAGKAESAASAST 336
Qy 259 AANNKTQATDASDOAAAADNTQAOTGDSFQSAQAQAVNNSDOESTTATEAQPSASSAS 318
Db 337 A---TTKAGEATEQASAAASASAKTSETN-----AKASETSAESSKTA-----AASAS 384
Qy 319 TAAVAANTSSANTYPAGQCTGWKSLAPWGVNGWGGQWAAASAAAAGYRVGSTPPSAGAV 378
Db 385 SAASSASSASAKDEATQASAAKSSATTAST---KATEAAGSATAAQAQSKSTAESAATR 441
Qy 379 A 379
Db 442 A 442

RESULT 8
US-10-873-528-109
; Sequence 109, Application US/10873528
; Publication No. US2005027681A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21129WO
; CURRENT APPLICATION NUMBER: US/10/873,528
; CURRENT FILING DATE: 2004-06-23
; PRIOR APPLICATION NUMBER: US/09/769,787
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 109
; LENGTH: 1236
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-873-528-109

Query Match      10.7%; Score 224.5; DB 6; Length 1236;
Best Local Similarity 21.5%; Pred. No. 8.3e-06;
Matches 84; Conservative 90; Mismatches 190; Indels 27; Gaps 7;

; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 3
; LENGTH: 971
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
US-11-052-554A-3

Query Match      10.8%; Score 227.5; DB 7; Length 971;
Best Local Similarity 25.5%; Pred. No. 4.2e-06;
Matches 92; Conservative 58; Mismatches 172; Indels 39; Gaps 8;

Qy 19 ATTLSAVKADDFDAQIASQDSKINNLTAAQOAAQAVNTTQGVSAIQVLAQAEQENQR 78
Db 121 AONTAAAKKSADASTSAREAAATATDAADSARAASTSAGQAASQASASSAGTASTKA 180
Qy 79 LEAQSATLGGQIQITLSSKIVARNESLKQARSQAKSNAATSYINAIINSKSVDAINRVS 138
Db 181 TEA-----SKSAAAESSKGAATSGA-AKTSETNAVVSQGSAA-----TS 221
Qy 139 AIREVVSANERKMLQOQBDKAAVEQKQOENQAAINTVAANQETIAQNTNALNTQQAQLEA 198
Db 222 ASTATTKASEAASARDASASKEAAKSSSETSAASSASSA-----ASSATAAGNSAKAAKT 276
Qy 199 AQLNLQALHTTAQDKATLVAKAAAEARQAAAAQAAAEAKAAAEKALQEQAAQAA 258
Db 277 SETNAKSETAAEQSAAAGSKTAAALSAASASTSAGQASASATAAGKAESAASAST 336
Qy 259 AANNKTQATDASDOAAAADNTQAOTGDSFQSAQAQAVNNSDOESTTATEAQPSASSAS 318
Db 337 A---TTKAGEATEQASAAASASAKTSETN-----AKASETSAESSKTA-----AASAS 384
Qy 319 TAAVAANTSSANTYPAGQCTGWKSLAPWGVNGWGGQWAAASAAAAGYRVGSTPPSAGAV 378
Db 385 SAASSASSASAKDEATQASAAKSSATTAST---KATEAAGSATAAQAQSKSTAESAATR 441
Qy 379 A 379
Db 442 A 442

RESULT 9
US-10-793-626-1682
; Sequence 1682, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1682
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1682

Query Match      10.2%; Score 215; DB 6; Length 149;
Best Local Similarity 39.7%; Pred. No. 2.6e-06;
Matches 50; Conservative 19; Mismatches 43; Indels 14; Gaps 5;

Qy 286 GDSTEQSAQAQAVNNSDOESTTATEAQPSASSASATAAANTSSANTYPAGQCTGWV-KSL 344
Db 20 GHTTHADAAE--NNNQOQST-----YNYSTTEVSF-SNSGNLTSQOCTWYVYDKT 67
Qy 345 APWGVNGWGGQWAAASAAAAGYRVGSTPSAGAVAVVNDGGYGHVAVVTGVQG-GQIQVQ 403
Db 68 GKGIGTGWGNSWATAQAAGFTVNTPEGAIMOSSEGAFGHVAFVESVNDGSIIVS 127
Qy 404 EANYAG 409
Db 128 EMNYDG 133

RESULT 10
US-10-793-626-2870
; Sequence 2870, Application US/10793626
```



Qy	300	--SDQESTTATEAQPSSASSASTAAVAANTSS-----ANTYPAGCCTGWV-----KSLAPWV	348
Db	127	IMPNQILTIPINGSGSGSGGTATQTSGNYTSPSFNQHLYTEGQCTWYVDFKRSQAKPI	186
Qy	349	GNVWGGGOWAAASAAAGYRVGTPPSAGAVAVMNDGGYGHVAVYVGQ-GQIQVQBNY	407
Db	187	STWSDAKYWNANDGYQVDNTPSVGAIMQSTPGPYGHVAVYVERINGDGSILISEMNY	246
Qy	408	AGNQSIGNYR	417
Db	247	ANGPYNNYR	256

RESULT 12  
 US-11-052-554A-211  
 ; Sequence 211, Application US/11052554A  
 ; Publication No. US20050288866A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Sachdeva, et al.  
 ; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN  
 ; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL  
 ; FILE REFERENCE: 30853/40359A  
 ; CURRENT APPLICATION NUMBER: US/11/052,554A  
 ; CURRENT FILING DATE: 2005-02-07  
 ; PRIOR APPLICATION NUMBER: US 60/589,227  
 ; PRIOR FILING DATE: 2004-07-20  
 ; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
 ; PRIOR FILING DATE: 2004-02-06  
 ; NUMBER OF SEQ ID NOS: 763  
 ; SOFTWARE: PatentIn version 3.3  
 ; SEQ ID NO 211  
 ; LENGTH: 1562  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus mutans UA159  
 US-11-052-554A-211

Query Match	9.7%	Score 204;	DB 7;	Length 1562;
Best Local Similarity	25.9%	Pred. No. 0.00016;		
Matches	85;	Conservative	62;	Mismatches 113;
			Indels	68;
			Gaps	12;

Qy	28	DDFDAQIASQDSKINLTAQQOAAQAVNTIQGVSAQTQQAELQENQRLAQSAT--	85
Db	151	DOYKSDVAHAEEAVAKIKAKNQKTEQ-----YEKDMAHKAEVERINAANAASK	200
Qy	86	-----LGQQIQTLSS--KTVARNESLKQOARSA-----OKSNAA--TSYINAII--	125
Db	201	TAYEAKLAQYQADLAAVOKTNAANQRAYQALAAAYQAEIKRVGEANNAKAAVDTAVAN	260
Qy	126	NSKSVSDAINRVSAIREVVSANEKMLQQQPCDKAAVEQKQENQAAINTVAAHQETIAQN	185
Db	261	NAKNT-----EIAAANEIEIKBNATAKAEYETKLAQYQAEIKRV-----QE	301
Qy	186	TNALNTQQAQLEAAQLNLQAEILTTAQDKATLVQAQKAAAEERARQAAAAQAABEAKAAAE	245
Db	302	ANAN-----EADYQAKLTAYQTELARVQKANADAKAAEAAVAANNAKNAALTAE	352
Qy	246	AKALQCAQAQAA--AAANNNTQATDASDQOAAADNT--QAAQTGDSGTEQSAQAQVNN	300
Db	353	NTAIKORNEAKATYEAALKQYEAADLAAVKKAANAANEADYQAKLTAYQTELARVQKAN--	410
Qy	301	DOESTTATEAQPSSASSASTAAVAANTSS	328
Db	411	ADAKAYEAVAANNAANALTAENTA	437

RESULT 13  
 US-10-793-626-3154  
 ; Sequence 3154, Application US/10793626  
 ; Publication No. US20050255478A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KIMMERLY, WILLIAM JOHN  
 ; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

```
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3154
; LENGTH: 1095
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3154

Query Match          9.3%; Score 195; DB 6; Length 1095;
Best Local Similarity 22.5%; Pred. No. 0.00034;
Matches 92; Conservative 61; Mismatches 165; Indels 90; Gaps 12;

Qy 13 GVTLSSTLSSAVKADDFDAQIASQDSKINNLTAAQQAQAQAVNTIQGVSAQTQQAEL 72
Db 340 GMTTDTANNYSKKREAE-EIQKAQOIINNGDATEQOITNETNRVNQAINAKNDL 398
Qy 73 QAENORLEAQSATLGGQIQITLSSKIVA-RNESLKQAQARSQAQKNAATSYINAINS 127
Db 399 RADKSLQENAYNQLIQNDTNGKKPASIQYQQAQAQIAETQYNNAKSEAHQILENSPSV 458
Qy 128 KVSVDAINRYSAIREVVSANEKMLQQOQEQDQKAAVEQKQENQA-----AINT 174
Db 459 NEVAQAQKQVEAVQLKVDNAIHMLQNKENSALVTAQNQLQQAQVNDQPLTTGTMQDSINN 518
Qy 175 VAA-----NQETIAQNTNA-----LNTQQAQLEAAQNLQ-----AELTTA 210
Db 519 YVAKRNEAQAIRNAEAVINNGDATAKQISDEKSKVEQALAHLDNAKQQLTADTTELOTA 578
Qy 211 QDQ-----KATLVAKAAEEAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 259
Db 579 VQQLNRGDTNNKPRINAYNKAIQSLEITQISAKNNANAVIQKPIRTVQEVNNALQOV 638
Qy 260 ANNTQTADSDQQAADN-----TQAQOTGDSQTEQS----- 292
Db 639 NQLNQQLTEAINGLQPLSNDLAKAARLNLENKINQTVQDGMTQQSIEAYQNAKRAQN 698
Qy 293 -----AAQAVNSD-----QESTATQAQPSASASTAAVAANTSSANTYP 333
Db 699 ESNATALINNGDAEQOITETDRVNVQQTNTLQTAI--NGLTVNKEP 744

RESULT 14
US-10-857-780-23
; Sequence 23, Application US/10857780
; Publication No. US20050272043A1
; GENERAL INFORMATION:
; APPLICANT: ROTH, RICHARD B.
; APPLICANT: BRAUN, ANDREAS
; APPLICANT: KAMMERER, STEFAN M.
; APPLICANT: NELSON, MATTHEW ROBERTS
; APPLICANT: RENELAND, RIKARD HENRY
; APPLICANT: HOVAL-WRIGHTSON, CAROLYN R.
; TITLE OF INVENTION: METHODS FOR IDENTIFYING RISK OF BREAST CANCER AND TREATMENTS
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: SEQ-4069-CP
; CURRENT APPLICATION NUMBER: US/10/857,780
; CURRENT FILING DATE: 2004-05-28
; PRIOR APPLICATION NUMBER: 10/723,681
; PRIOR FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: 60/490,234
; PRIOR FILING DATE: 2003-07-24
; PRIOR APPLICATION NUMBER: 60/525,239
; PRIOR FILING DATE: 2003-11-25
; NUMBER OF SEQ ID NOS: 4962
; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 23
; LENGTH: 2101
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-857-780-23

Query Match          9.2%; Score 194.5; DB 6; Length 2101;
Best Local Similarity 22.4%; Pred. No. 0.00076;
Matches 95; Conservative 62; Mismatches 181; Indels 87; Gaps 14;

Qy 2 KKRILSAVLVSGVTLSSTLSSA-----VKADDFDAQIASQDSKI-----NNLTAAQQA 51
Db 405 KGEVLGDLVQLQLETLKQEAATLAANNLTQQAQVEMLETERGQEQAKLALAEGRHFEEKQOL 464
Qy 52 QAQVNTIQGVSAQTQQAEL-----QAENORLEAQSATLGGQIQITLSSKIVARNE---SL 104
Db 465 SSILITDLQSSISNLSQAKEELEQAQAHGARLTAAQVASLTSETLTNLTATQQQDQELAGL 524
Qy 105 KQQAARSQAQKNAATSYINAINSVSVDAINRYSAIREVVSANEKMLQQOQEQDQKAAVEQK 164
Db 525 KQQAQKQQAQLAQT-----LQQQEQASQGLRHQVEQLSSSLKQKEQQLKEVAEK 573
Qy 165 Q-----QENQAQINVTAAQNTIAQNTNALNTQQAQLE---AAQLNL----- 203
Db 574 QEATRDQHAQQLATAAEEEREASLRERDAALKQLEALEKEKAQKLEILQQQLQVANEARDS 633
Qy 204 -QAEITTAQDQKATL---VAQKAAAEAAQAAAAQAAAAAQAQAQAQAQAQAQAQAQAQA 259
Db 634 AQTSTVTAQREKAELSRKVEELQACVETARQEQHEAQAQAQVAELEQLRSEQOKATEKERV 693
Qy 260 ANNTQ-----ATDASDQAAAAADNTQAQOTGDSQTEQSAA 294
Db 694 AQEKDQLQEQQLAKSLKVTKSLSEKERRADALEEQORCISELK-AETRSILVEQHKR 752
Qy 295 QAVN-NSDQESTTATBAQ-----PSASSASTAAV---AANTSSANTYPAGQCTGVKSLAP 346
Db 753 ERKELEERAGRKGLEARLLQLGEAQAETEVLRRELAEMAAQAHTAESCEQLVKEVAA 812
Qy 347 WVGNY 351
Db 813 WRDGY 817

RESULT 15
US-10-485-517-212
; Sequence 212, Application US/10485517
; Publication No. US2005025629A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629W0
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 212
; LENGTH: 1448
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-212

Query Match          9.1%; Score 192.5; DB 6; Length 1448;
Best Local Similarity 25.0%; Pred. No. 0.00065;
Matches 92; Conservative 60; Mismatches 147; Indels 69; Gaps 16;

Qy 7 SAVLVSGVTLSSTLSSAVK-----ADDFDAQIASQDSKINNLTAAQQAQAQV---- 55
```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:15:51 ; Search time 0.910282 Seconds  
(without alignments)  
302.734 Million cell updates/sec

Title: US-10-797-821-24

Perfect score: 104

Sequence: 1 DANFDSIRVDVNDVADLLQ 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US03\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	39.4	295	6	US-10-467-657-5222
2	40.5	38.9	449	6	US-10-763-712A-73
3	40.5	38.9	772	6	US-10-858-730-77
4	40	38.5	430	6	US-10-467-657-2346
5	40	38.5	480	7	US-11-136-244-9
6	40	38.5	619	6	US-10-999-886-3
7	40	38.5	619	7	US-11-136-244-3
8	40	38.5	640	6	US-10-999-886-4
9	40	38.5	640	7	US-11-136-244-4
10	39	37.5	199	6	US-10-793-626-2546
11	39	37.5	447	6	US-10-967-527A-14
12	39	37.5	448	6	US-10-967-527A-16
13	39	37.5	610	7	US-11-055-822-1146
14	39	37.5	610	7	US-11-124-291-6
15	39	37.5	737	7	US-11-128-420-9
16	39	37.5	884	6	US-10-878-556A-9
17	39	37.5	974	6	US-10-821-234-1152
18	38.5	37.0	503	7	US-11-078-391-1
19	38	36.5	327	6	US-10-510-386-34
20	38	36.5	325	7	US-11-074-176-370
21	38	36.5	332	6	US-10-793-626-2456
22	38	36.5	505	6	US-10-793-626-2220
23	38	36.5	1531	7	US-11-087-227-24
24	38	36.5	1531	7	US-11-186-284-211
25	37.5	36.1	359	6	US-10-467-657-1676

Sequence 2, Appli  
Sequence 4, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 8534, Ap  
Sequence 283, App  
Sequence 281, App  
Sequence 1116, Ap  
Sequence 2832, Ap  
Sequence 1473, Ap  
Sequence 63, Appl  
Sequence 105, App  
Sequence 278, App  
Sequence 4306, Ap  
Sequence 132, App  
Sequence 963, App  
Sequence 226, App  
Sequence 32, Appl

26 37 35.6 485 6 US-10-630-203-2  
27 37 35.6 485 6 US-10-630-203-4  
28 37 35.6 485 7 US-11-103-037-1  
29 37 35.6 485 7 US-11-103-037-2  
30 37 35.6 485 7 US-11-195-538-1  
31 37 35.6 485 7 US-11-195-538-2  
32 37 35.6 856 6 US-10-467-657-8534  
33 37 35.6 1461 7 US-11-052-554A-283  
34 37 35.6 5291 7 US-11-052-554A-281  
35 36.5 35.1 320 7 US-11-055-822-1116  
36 36.5 35.1 584 6 US-10-793-626-2832  
37 36.5 35.1 1058 6 US-10-821-234-1473  
38 36.5 35.1 1058 6 US-10-878-556A-63  
39 36.5 35.1 1058 7 US-11-089-642-105  
40 36 34.6 181 6 US-10-467-657-278  
41 36 34.6 181 6 US-10-467-657-4306  
42 36 34.6 208 7 US-11-156-084-132  
43 36 34.6 316 6 US-10-995-561-963  
44 36 34.6 329 7 US-11-082-389-226  
45 36 34.6 330 7 US-11-125-402-32

#### ALIGNMENTS

RESULT 1  
US-10-467-657-5222  
; Sequence 5222, Application US/10467657  
; Publication No. US20050260581A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; APPLICANT: FONTANA Maria Rita  
; APPLICANT: PIZZA Mariagrazia  
; APPLICANT: MASIGNANI Vega  
; APPLICANT: MONACI Elisabetta  
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/467,657  
; CURRENT FILING DATE: 2003-08-11  
; PRIOR APPLICATION NUMBER: GB-0103424.8  
; PRIOR FILING DATE: 2001-02-12  
; NUMBER OF SEQ ID NOS: 9218  
; SOFTWARE: Seqwin99, version 1.04  
; SEQ ID NO 5222  
; LENGTH: 295  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-5222

Query Match 39.4%; Score 41; DB 6; Length 295;  
Best Local Similarity 19.0%; Pred. No. 22;  
Matches 4; Conservative 14; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVADLLQ 21  
Db 179 NSNPDAVRLDKIEIHNEIIE 199

RESULT 2  
US-10-763-712A-73  
; Sequence 73, Application US/10763712A  
; Publication No. US20050266541A1  
; GENERAL INFORMATION:  
; APPLICANT: Solazyme, Inc.  
; APPLICANT: Dillion, Harrison F.  
; TITLE OF INVENTION: Methods and Compositions for Evolving Microbial Hydrogen  
; TITLE OF INVENTION: Production  
; FILE REFERENCE: H2042101-CIP  
; CURRENT APPLICATION NUMBER: US/10/763,712A  
; CURRENT FILING DATE: 2004-01-21  
; PRIOR APPLICATION NUMBER: US 10/287,750  
; PRIOR FILING DATE: 2002-11-04  
; PRIOR APPLICATION NUMBER: US 10/411,910

```
; PRIOR FILING DATE: 2003-04-12
; PRIOR APPLICATION NUMBER: US 60/500,032
; PRIOR FILING DATE: 2003-09-03
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 73
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Clostridium perfringens
US-10-763-712A-73

Query Match      38.9%; Score 40.5; DB 6; Length 449;
Best Local Similarity 42.9%; Pred. No. 44;
Matches 9; Conservative 4; Mismatches 5; Indels 3; Gaps 1;

QY      4 FDSIRVDVAVN---VDADLLQ 21
Db      101 FDAITDPIDNSKYIDPNLCQ 121
|||||:|:|

RESULT 3
US-10-858-730-77
; Sequence 77, Application US/10858730
; Publication No. US20050255568A1
; GENERAL INFORMATION:
; APPLICANT: Bailey, Richard B.
; APPLICANT: Blomquist, Paul
; APPLICANT: Doten, Reed
; APPLICANT: Driggers, Edward M.
; APPLICANT: Madden, Kevin T.
; APPLICANT: O'Leary, Jessica
; APPLICANT: O'Toole, George
; APPLICANT: Trueheart, Joshua
; APPLICANT: Walbridge, Michael J.
; APPLICANT: Yorgey, Peter S.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID
; FILE REFERENCE: 14184-030001
; CURRENT APPLICATION NUMBER: US/10/858,730
; CURRENT FILING DATE: 2004-06-01
; PRIOR APPLICATION NUMBER: US 60/475,000
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/551,860
; PRIOR FILING DATE: 2004-03-10
; NUMBER OF SEQ ID NOS: 364
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 77
; LENGTH: 772
; TYPE: PRT
; ORGANISM: Streptomyces coelicolor
US-10-858-730-77

Query Match      38.9%; Score 40.5; DB 6; Length 772;
Best Local Similarity 47.4%; Pred. No. 85;
Matches 9; Conservative 5; Mismatches 4; Indels 1; Gaps 1;

QY      2 ANFDSIRVDVAVNVDADLL 20
Db      660 AEFGDI-VQAIDDLADAVI 677
|||||:|:|:|:|:|:|:|

RESULT 4
US-10-467-657-2346
; Sequence 2346, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:

; PRIOR FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 2346
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-2346

Query Match      38.5%; Score 40; DB 6; Length 430;
Best Local Similarity 63.6%; Pred. No. 50;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      11 AVDNVDADLLQ 21
Db      83 AVDNIDLDIYQ 93
|||||:|:|

RESULT 5
US-11-136-244-9
; Sequence 9, Application US/11136244
; Publication No. US20060003408A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Fiske, Susan M.
; APPLICANT: Lantz, Suzanne E.
; APPLICANT: Neefe-Kruithof, Paulien
; APPLICANT: Pepsin, Michael Jay
; APPLICANT: Jayarama, Shetty K.
; TITLE OF INVENTION: Acid-Stable Alpha Amylases Having Granular Starch Hydrolyzing
; TITLE OF INVENTION: Activity and Enzyme Compositions
; FILE REFERENCE: GC873
; CURRENT APPLICATION NUMBER: US/11/136,244
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US 60/575,175
; PRIOR FILING DATE: 2004-05-27
; PRIOR APPLICATION NUMBER: US 60/605,437
; PRIOR FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: US 60/647,925
; PRIOR FILING DATE: 2005-01-28
; PRIOR APPLICATION NUMBER: PCT/US04/40040
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: PCT/US04/41276
; PRIOR FILING DATE: 2004-12-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Aspergillus kawachi
US-11-136-244-9

Query Match      38.5%; Score 40; DB 7; Length 480;
Best Local Similarity 42.9%; Pred. No. 57;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY      5 DSIRVDVAVNVDAD 18
Db      201 DGLRIDSVEVBPD 214
|||||:|:|:|:|

RESULT 6
US-10-999-886-3
; Sequence 3, Application US/10999886
; Publication No. US20050266543A1
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Lantero, Oreste
; APPLICANT: Shetty, Jayarama K.
```

; APPLICANT: Suzanne, Lantz E.  
 ; APPLICANT: Michael, Pepsin J.  
 ; TITLE OF INVENTION: Heterologous Expression of an Aspergillus kawachi Acid-Stable  
 ; TITLE OF INVENTION: Alpha Amylase and Applications in Granular Starch Hydrolysis  
 ; FILE REFERENCE: GC841  
 ; CURRENT APPLICATION NUMBER: US/10/999,886  
 ; CURRENT FILING DATE: 2004-11-30  
 ; PRIOR APPLICATION NUMBER: US 60/605,437  
 ; PRIOR FILING DATE: 2004-08-30  
 ; PRIOR APPLICATION NUMBER: US 60/575,175  
 ; PRIOR FILING DATE: 2004-05-27  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 3  
 ; LENGTH: 619  
 ; TYPE: PRT  
 ; ORGANISM: Aspergillus kawachi  
 US-10-999-886-3

Query Match 38.5%; Score 40; DB 6; Length 619;  
 Best Local Similarity 42.9%; Pred. No. 77;  
 Matches 6; Conservative 5; Mismatches 3; Indels 3; Gaps 0;

Qy 5 DSIRVDVNDVAD 18

Db 201 DGLRIDSVEEVPD 214

RESULT 7  
 US-11-136-244-3  
 ; Sequence 3, Application US/11136244  
 ; Publication No. US20060003408A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dunn-Coleman, Nigel  
 ; APPLICANT: Fiske, Susan M.  
 ; APPLICANT: Lantz, Suzanne E.  
 ; APPLICANT: Neefe-Kruitthof, Paulien  
 ; APPLICANT: Pepsin, Michael Jay  
 ; APPLICANT: Jayarama, Shetty K.  
 ; TITLE OF INVENTION: Acid-Stable Alpha Amylases Having Granular Starch Hydrolyzing  
 ; TITLE OF INVENTION: Activity and Enzyme Compositions  
 ; FILE REFERENCE: GC873  
 ; CURRENT APPLICATION NUMBER: US/11/136,244  
 ; CURRENT FILING DATE: 2005-05-24  
 ; PRIOR APPLICATION NUMBER: US 60/575,175  
 ; PRIOR FILING DATE: 2004-05-27  
 ; PRIOR APPLICATION NUMBER: US 60/605,437  
 ; PRIOR FILING DATE: 2004-08-30  
 ; PRIOR APPLICATION NUMBER: US 60/647,925  
 ; PRIOR FILING DATE: 2005-01-28  
 ; PRIOR APPLICATION NUMBER: PCT/US04/40040  
 ; PRIOR FILING DATE: 2004-11-30  
 ; PRIOR APPLICATION NUMBER: PCT/US04/41276  
 ; PRIOR FILING DATE: 2004-12-09  
 ; NUMBER OF SEQ ID NOS: 13  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 3  
 ; LENGTH: 619  
 ; TYPE: PRT  
 ; ORGANISM: Aspergillus kawachi  
 US-11-136-244-3

Query Match 38.5%; Score 40; DB 7; Length 619;  
 Best Local Similarity 42.9%; Pred. No. 77;  
 Matches 6; Conservative 5; Mismatches 3; Indels 3; Gaps 0;

Qy 5 DSIRVDVNDVAD 18

Db 201 DGLRIDSVEEVPD 214

RESULT 8  
 US-10-999-886-4

; Sequence 4, Application US/10999886  
 ; Publication No. US20050286543A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Genencor International, Inc.  
 ; APPLICANT: Dunn-Coleman, Nigel  
 ; APPLICANT: Lantero, Orsette  
 ; APPLICANT: Shetty, Jayarama K.  
 ; APPLICANT: Suzanne, Lantz E.  
 ; APPLICANT: Michael, Pepsin J.  
 ; TITLE OF INVENTION: Heterologous Expression of an Aspergillus kawachi Acid-Stable  
 ; TITLE OF INVENTION: Alpha Amylase and Applications in Granular Starch Hydrolysis  
 ; FILE REFERENCE: GC841  
 ; CURRENT APPLICATION NUMBER: US/10/999,886  
 ; CURRENT FILING DATE: 2004-11-30  
 ; PRIOR APPLICATION NUMBER: US 60/605,437  
 ; PRIOR FILING DATE: 2004-08-30  
 ; PRIOR APPLICATION NUMBER: US 60/575,175  
 ; PRIOR FILING DATE: 2004-05-27  
 ; NUMBER OF SEQ ID NOS: 7  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 4  
 ; LENGTH: 640  
 ; TYPE: PRT  
 ; ORGANISM: Aspergillus kawachi  
 US-10-999-886-4

Query Match 38.5%; Score 40; DB 6; Length 640;  
 Best Local Similarity 42.9%; Pred. No. 81;  
 Matches 6; Conservative 5; Mismatches 3; Indels 3; Gaps 0;

Qy 5 DSIRVDVNDVAD 18

Db 222 DGLRIDSVEEVPD 235

RESULT 9  
 US-11-136-244-4  
 ; Sequence 4, Application US/11136244  
 ; Publication No. US20060003408A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dunn-Coleman, Nigel  
 ; APPLICANT: Fiske, Susan M.  
 ; APPLICANT: Lantz, Suzanne E.  
 ; APPLICANT: Neefe-Kruitthof, Paulien  
 ; APPLICANT: Pepsin, Michael Jay  
 ; APPLICANT: Jayarama, Shetty K.  
 ; TITLE OF INVENTION: Acid-Stable Alpha Amylases Having Granular Starch Hydrolyzing  
 ; TITLE OF INVENTION: Activity and Enzyme Compositions  
 ; FILE REFERENCE: GC873  
 ; CURRENT APPLICATION NUMBER: US/11/136,244  
 ; CURRENT FILING DATE: 2005-05-24  
 ; PRIOR APPLICATION NUMBER: US 60/575,175  
 ; PRIOR FILING DATE: 2004-05-27  
 ; PRIOR APPLICATION NUMBER: US 60/605,437  
 ; PRIOR FILING DATE: 2004-08-30  
 ; PRIOR APPLICATION NUMBER: US 60/647,925  
 ; PRIOR FILING DATE: 2005-01-28  
 ; PRIOR APPLICATION NUMBER: PCT/US04/40040  
 ; PRIOR FILING DATE: 2004-11-30  
 ; PRIOR APPLICATION NUMBER: PCT/US04/41276  
 ; PRIOR FILING DATE: 2004-12-09  
 ; NUMBER OF SEQ ID NOS: 13  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 4  
 ; LENGTH: 640  
 ; TYPE: PRT  
 ; ORGANISM: Aspergillus kawachi  
 US-11-136-244-4

Query Match 38.5%; Score 40; DB 7; Length 640;  
 Best Local Similarity 42.9%; Pred. No. 81;  
 Matches 6; Conservative 5; Mismatches 3; Indels 3; Gaps 0;

```
QY      5 DSIRVDVAVNVAD 18
      ||:|:|:|:|:|
Db      222 DGLRIDSVVEVPD 235

RESULT 10
US-10-793-626-2546
; Sequence 2546, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2546
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2546

Query Match      37.5%; Score 39; DB 6; Length 199;
Best Local Similarity 53.8%; Pred. No. 28;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      9 VDAVDNVVDADLLQ 21
      |||:|:|:|:|:|
Db      167 IDAVENDDASILR 179

RESULT 11
US-10-967-527A-14
; Sequence 14, Application US/10967527A
; Publication No. US20050256041A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Holloway, James L.
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: Ztnfr14, A Tumor Necrosis Factor
; FILE REFERENCE: 03-17
; CURRENT APPLICATION NUMBER: US/10/967,527A
; CURRENT FILING DATE: 2004-10-18
; PRIOR APPLICATION NUMBER: 60/511,698
; PRIOR FILING DATE: 2003-10-16
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 447
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-967-527A-14

Query Match      37.5%; Score 39; DB 6; Length 447;
Best Local Similarity 50.0%; Pred. No. 74;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY      8 RVDADVNDVADLLQ 21
      ||:|:|:|:|:|
Db      419 RLDVAVSLCADILE 432

RESULT 12
US-10-967-527A-16
; Sequence 16, Application US/10967527A
; Publication No. US20050256041A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Holloway, James L.
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: Ztnfr14, A Tumor Necrosis Factor
; FILE REFERENCE: 03-17
; CURRENT APPLICATION NUMBER: US/10/967,527A
; CURRENT FILING DATE: 2004-10-18
; PRIOR APPLICATION NUMBER: 60/511,698
; PRIOR FILING DATE: 2003-10-16
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 448
; TYPE: PRT
; ORGANISM: mus musculus
US-10-967-527A-16

Query Match      37.5%; Score 39; DB 6; Length 448;
Best Local Similarity 50.0%; Pred. No. 75;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY      8 RVDADVNDVADLLQ 21
      ||:|:|:|:|:|
Db      420 RLDVAVSLCADILE 433

RESULT 13
US-11-055-822-1146
; Sequence 1146, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; FILE REFERENCE: EGI-121CPCN
; CURRENT APPLICATION NUMBER: US/11/055,822
; CURRENT FILING DATE: 2005-02-11
; PRIOR APPLICATION NUMBER: 09/606,740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141,031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142,101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148,613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187,970
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19930476.9
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19931415.2
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931418.7
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931419.5
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1158
; SEQ ID NO 1146
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-11-055-822-1146

Query Match      37.5%; Score 39; DB 7; Length 610;
Best Local Similarity 43.8%; Pred. No. 1.1e+02;
```



Matches	7;	Conservative	5;	Mismatches	4;	Indels	0;	Gaps	0;
---------	----	--------------	----	------------	----	--------	----	------	----

Qy 1 DANFDSIRVDAVDNVD 16  
| : | : | : | : | : | : |  
Db 276 DFHV DGLRLDAVHSLD 291

Qy 9 VDAVDNVADLL 20  
|||:|  
Db 672 VDMEDNLDQDLL 683

Search completed: February 10, 2006, 23:26:00  
Job time : 1.91028 secs

```

RESULT 14
US-11-124-291-6
; Sequence 6, Application US/11124291
; Publication No. US20050266536A1
; GENERAL INFORMATION:
; APPLICANT: WOLF, ANDREAS
; APPLICANT: SCHISCHKA, NATALIE
; APPLICANT: HERMANN, THOMAS
; APPLICANT: MORBACH, SUSANNE
; APPLICANT: KRAEMER, REINHARD
; TITLE OF INVENTION: PROCESS FOR THE FERMENTATIVE PREPARATION OF L-AMINO ACIDS USING
; TITLE OF INVENTION: CORYNEFORM BACTERIA
; FILE REFERENCE: 223168US0X
; CURRENT APPLICATION NUMBER: US/11/124,291
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US/10/212,219
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US 60/316,276
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: DE 101 39 062.9
; PRIOR FILING DATE: 2001-09-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-11-124-291-6

```

```
Query Match          37.5%; Score 39; DB 7; Length 610;
Best Local Similarity 43.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
```

Qy	1	DANFDSIRVDAVDNVD	16
		:   :   :   :	
Db	276	DFHVDGLRLDAVHSLD	291

```

RESULT 15
US-11-128-420-9
; Sequence 9, Application US/11128420
; Publication No. US20050266463A1
; GENERAL INFORMATION:
; APPLICANT: Nicolaides, Nicholas
; APPLICANT: Grasso, Luigi
; APPLICANT: Sasse, Philip
; APPLICANT: Kinzler, Kenneth
; APPLICANT: Vogelstein, Bert
; TITLE OF INVENTION: A method for generating hypermutable
; TITLE OF INVENTION: plants
; FILE REFERENCE: 01107.00069
; CURRENT APPLICATION NUMBER: US/11/128,420
; CURRENT FILING DATE: 2005-05-13
; PRIOR APPLICATION NUMBER: 60/183,333
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 9
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-11-128-420-9

```

Query Match 37.5%; Score 39; DB 7; Length 737;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

五

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run-on: February 10, 2006, 22:05:33 ; Search time 15.9234 Seconds  
(without alignments)  
607.053 Million cell updates/sec

Title: US-10-797-821-25

Perfect score: 108

Sequence: 1 DANFDSIRVDVNDVADLLQI 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_21.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	100.0	22	2 AAR43694	Aar43694 GT subseq
2	108	100.0	22	2 AAW34158	Aaw34158 GTF anti
3	108	100.0	22	7 ADD93645	Add93645 Streptoco
4	108	100.0	22	9 ADX37268	Adx37268 Streptoco
5	108	100.0	24	2 AAW34163	Aaw34163 GTF anti
6	108	100.0	1017	5 AAU79285	Aau79285 Streptoco
7	108	100.0	1375	5 AAU98028	Aau98028 S. mutans
8	108	100.0	1375	5 AAU79288	Aau79288 Streptoco
9	108	100.0	1375	7 ADD93655	Add93655 Streptoco
10	108	100.0	1375	9 ADX37278	Adx37278 Streptoco
11	108	100.0	1475	5 AAU98036	Aau98036 S. mutans
12	108	100.0	1475	5 AAU98037	Aau98037 S. mutans
13	108	100.0	1475	5 AAU98040	Aau98040 S. mutans
14	108	100.0	1475	5 AAU98033	Aau98033 S. mutans
15	108	100.0	1475	5 AAU98039	Aau98039 S. mutans
16	108	100.0	1475	5 AAU98027	Aau98027 S. mutans
17	108	100.0	1475	5 AAU98032	Aau98032 S. mutans
18	108	100.0	1475	7 ADD93654	Add93654 Streptoco
19	108	100.0	1475	9 ADX37277	Adx37277 Streptoco
20	108	100.0	1476	5 AAU79284	Aau79284 Streptoco
21	108	100.0	1499	7 ADC54806	Adc54806 Protein S
22	108	100.0	1590	7 ADD93657	Add93657 Streptoco
23	108	100.0	1590	9 ADX37280	Adx37280 Streptoco
24	107	99.1	1475	5 AAU98030	Aau98030 S. mutans

ALIGNMENTS

RESULT 1

AAR43694

ID AAR43694 standard; peptide; 22 AA.

XX AAR43694;

AC AAR43694;

XX 25-MAR-2003 (revised)

DT 20-MAY-1994 (first entry)

XX GT subsequence.

XX GT; glucosyltransferase; vaccine; T-cell; B-cell; reaction;

KW immunoreponse; peptidyl core matrix; dental caries; diptheria; tetanus;

KW measles; polio.

XX Synthetic.

OS WO9322341-A1.

PN 11-NOV-1993.

XX 30-APR-1993; 93WO-US004094.

XX 01-MAY-1992; 92US-00877295.

XX (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.

PI Taubman MA, Smith DJ;

XX WPI; 1993-368721/46.

XX Synthetic glucosyl-transferase peptide(s) - provoke T- and B-cell

XX reactions in mammals, and are effective against dental caries.

XX Claim 3; Page 23; 38pp; English.

PS The sequences (AAR43694-98) are subsequences from GT, they provoke T- and  
CC B-cell responses in mammals. The can be used to produce a vaccine  
CC comprising 2 of these sequences attached to a peptidyl core matrix. They  
CC are useful in producing T-cell responses to the GT enzyme causing a  
CC reduction of colonisation or accumulation of mutans streptococcal strains  
CC in mammals. The vaccines can be used in preventing dental caries.  
CC (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 22 AA;

Query Match 100.0%; Score 108; DB 2; Length 22;

Best Local Similarity 100.0%; Pred. No. 4.6e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 DANFDSIRVDVNDVADLLQI 22

RESULT 2  
AAW34158  
ID AAW34158 standard; peptide; 22 AA.  
XX AC AAW34158;  
XX AC  
XX DT 25-MAR-2003 (revised)  
XX DT 18-FEB-1998 (first entry)  
XX DE  
XX DE GTF antigenic peptide #3.  
XX KW Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;  
XX KW surface domain; glucan-binding domain; mutans streptococcal strain;  
XX KW immune response; glucan-binding activity; dental caries prevention.  
XX OS Streptococcus mutans.  
XX PN US5686075-A.  
XX PD 11-NOV-1997.  
XX PF 30-APR-1993; 93US-00057162.  
XX PR 01-MAY-1992; 92US-00877295.  
XX PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.  
XX PI Smith DJ, Taubman MA;  
XX WPI; 1997-558089/51.  
XX Immunogenic compositions containing streptococcal glucosyltransferase  
PT peptide(s) - used for provoking immune response to streptococcal  
PT glucosyltransferase for preventing dental caries.  
XX Claim 2; Col 13; 11pp; English.  
XX CC AAW34156-W34160 represent immunogenic fragments of the Streptomyces  
CC mutans glucosyltransferase (GTF) enzyme. AAW34157 and AAW34158 are from  
CC the catalytic domain of GTF, while AAW34160 is from the GTF surface  
CC domain. AAW34156 and AAW34159 are from the glucan-binding domain of GTF.  
CC These sequences can all be used in the immunogenic composition of the  
CC invention. The composition of the invention can alternatively comprise  
CC one or more of these sequences linked to a lysine core matrix (see  
CC AAW34161-W34165). A composition comprising one of these sequences can be  
CC administered to a mammal to raise an immune response, in a method for  
CC interfering with the enzymatic activity of streptococcal  
CC glucosyltransferase in a mammal. The immune response results in reduction  
CC of the colonisation or accumulation of mutans streptococcal strains in  
CC the mammal. Compositions containing AAW34156 specifically interfere with  
CC the glucan-binding activity of the streptococcal glucosyltransferase. The  
CC peptides can also be used in vaccines for preventing dental caries in  
CC mammals. (Updated on 25-MAR-2003 to correct PF field.)  
XX  
SQ Sequence 22 AA;

Query Match 100.0%; Score 108; DB 2; Length 22;  
Best Local Similarity 100.0%; Pred. No. 4.6e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 DANFDSIRVDVNDVADLLQI 22

RESULT 3  
ADD93645  
ID ADD93645 standard; peptide; 22 AA.  
XX AC ADD93645;  
XX DT 29-JAN-2004 (first entry)  
XX DE Streptococcus mutans glucosyltransferase-B catalytic domain peptide.  
XX KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.  
XX OS Streptococcus mutans.  
XX PN WO2003075845-A2.  
XX PD 18-SEP-2003.  
XX PF 07-MAR-2003; 2003WO-US006962.  
XX PR 07-MAR-2002; 2002US-0363209P.  
XX PR 08-AUG-2002; 2002US-0402483P.  
XX PA (FORS-) FORSYTH INST.  
XX PI Smith DJ, Taubman MA;  
XX WPI; 2003-845091/78.  
XX Composition useful as vaccines for dental caries comprises a fragment of  
PT a glucan binding protein-B binding to a major histocompatibility complex  
PT class II protein.  
XX Claim 17; Page 17; 49pp; English.  
XX CC The present sequence is that of a catalytic domain peptide fragment of  
CC Streptococcus mutans glucosyltransferase. The peptide can be used in  
CC immunogenic compositions and subunit vaccines for dental caries. These  
CC compositions comprise a major histocompatibility complex (MHC) class II  
CC protein-binding peptide from Streptococcus mutans glucan binding protein  
CC -B (GbpB) covalently linked with a (preferably catalytic domain) peptide  
CC subunit of a streptococcal glucosyltransferase. The compositions are used  
CC in a claimed method of eliciting production of an antibody in a mammal.  
CC Dieptopic or multi-epitopic polypeptides can be prepared synthetically or  
CC by recombinant DNA technology. Antibodies raised against MHC class II  
CC binding fragments of GbpB can be used in passive immunisation.  
XX  
SQ Sequence 22 AA;

Query Match 100.0%; Score 108; DB 7; Length 22;  
Best Local Similarity 100.0%; Pred. No. 4.6e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 DANFDSIRVDVNDVADLLQI 22

RESULT 4  
ADX37268  
ID ADX37268 standard; peptide; 22 AA.  
XX AC ADX37268;  
XX DT 21-APR-2005 (first entry)  
XX DE Streptococcus mutant glucan binding protein B peptide #25.  
XX KW immunogenicity; immune stimulation; glucan binding protein-B;  
XX KW microparticle; major histocompatibility complex; tooth disease.  
XX OS Streptococcus mutans.  
XX

```

PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
XX 09-MAR-2004; 2004US-00797821.
XX
XX 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
XX Smith DJ, Taubman MA;
XX
XX WPI; 2005-151644/16.
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX Claim 8; SEQ ID NO 25; 73pp; English.
XX
XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB-derived peptide of the invention.
XX
SQ Sequence 22 AA;
XX
Query Match 100.0%; Score 108; DB 9; Length 22;
Best Local Similarity 100.0%; Pred. NO. 4.6e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 DANFDSIRVDVNDVDADLLQI 22
Db |||||||||||||||||||
1 DANFDSIRVDVNDVDADLLQI 22

RESULT 5
AAW34163
ID AAW34163 standard; peptide; 24 AA.
XX
AC AAW34163;
XX
DT 25-MAR-2003 (revised)
DT 18-FEB-1998 (first entry)
XX
XX GTF antigenic peptide #3 linked to polylysine core.
XX
XX Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;
KW surface domain; glucan-binding domain; mutans streptococcal strain;
KW immune response; glucan-binding activity; dental caries prevention.
XX
XX Synthetic.
OS Streptococcus mutans.
XX
XX Key Location/Qualifiers
XX Peptide 1..22
FT /label= GTF antigenic peptide #3 (see AAW34158)
FT /note= "attached to the dendritic polylysine core via the
FT alpha-amino group of Lys(23); a second copy of the
FT antigenic 22-mer is linked to Lys(23) via the omega amino
FT group"
FT Modified-site 23
FT /note= "Lys(23) is linked to one copy of the antigenic
FT peptide through the alpha-amino group, and to a second
FT

```

---

```

FT copy of the peptide (not shown) via the omega amino
FT group"
FT Modified-site 24
FT /note= "the alpha amino acid group of Lys(24) forms a
FT peptide linkage with the carboxyl amino group of Lys(23);
FT the omega amino group of Lys(24) forms a peptide bond
FT with a second Lys residue analogous to Lys(23)"
XX
XX US5686075-A.
XX
XX 11-NOV-1997.
XX
XX 30-APR-1993; 93US-00057162.
XX
XX 01-MAY-1992; 92US-00877295.
XX
XX (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
XX
XX Smith DJ, Taubman MA;
XX
XX WPI; 1997-558089/51.
XX
XX Immunogenic compositions containing streptococcal glucosyltransferase
PT peptide(s) - used for provoking immune response to streptococcal
PT glucosyltransferase for preventing dental caries.
XX
XX Claim 12; Col; 11pp; English.
XX
XX AAW34161-W34165 represent the Streptomyces mutans glucosyltransferase
CC (GTF) enzyme immunogenic fragments shown in AAW34156-W34160 linked to a
CC polylysine core. AAW34157 and AAW34158 are from the catalytic domain of
CC GTF, while AAW34160 is from the GTF surface domain. AAW34156 and AAW34159
CC are from the glucan-binding domain of GTF. These sequences, and the
CC immunogenic fragments shown in AAW34156-W34160 can all be used in the
CC immunogenic composition of the invention. A composition comprising one of
CC these sequences can be administered to a mammal to raise an immune
CC response, in a method for interfering with the enzymatic activity of
CC streptococcal glucosyltransferase in a mammal. The immune response
CC results in reduction of the colonisation or accumulation of mutans
CC streptococcal strains in the mammal. Compositions containing AAW34156
CC specifically interfere with the glucan-binding activity of the
CC streptococcal glucosyltransferase. The peptides can also be used in
CC vaccines for preventing dental caries in mammals. (Updated on 25-MAR-2003
CC to correct PF field.)
XX
XX Sequence 24 AA;
XX
Query Match 100.0%; Score 108; DB 2; Length 24;
Best Local Similarity 100.0%; Pred. NO. 5.2e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 DANFDSIRVDVNDVDADLLQI 22
Db |||||||||||||||||||
1 DANFDSIRVDVNDVDADLLQI 22

RESULT 6
AAU79285
ID AAU79285 standard; protein; 1017 AA.
XX
AC AAU79285;
XX
XX 13-AUG-2002 (first entry)
XX
XX Streptococcus mutans monoclonal antibody-related protein #2.
XX
XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;
KW glucosyl transferase-B; immunotherapy.
XX
XX Streptococcus mutans.
OS
XX
XX JP2002114709-A.
XX

```



XX 04-OCT-2000; 2000JP-00304889.  
XX  
XX 04-OCT-2000; 2000JP-00304889.  
XX  
XX (UYN-1) UNIV NIPPON.  
XX  
XX WPI; 2002-448885/48.  
XX  
XX Anti-carries agent composed of a monoclonal antibody against an inhibitory  
PT enzyme against water insoluble glucan synthetase of Streptococcus-  
PT B (GTF-B) of Streptococcus mutans.  
XX  
XX Disclosure; Page 22-25; 28pp; Japanese.  
XX  
XX The invention relates to a monoclonal antibody against dental caries and  
CC an anti-carries agent composed of a monoclonal antibody produced by  
CC Streptococcus mutans, particularly mouse-hybridoma MHP136 (FERM P-17566)  
CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having  
CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
CC transferase-B. The monoclonal antibody specifically inhibits water  
CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl  
CC transferase-B and is used in the immunotherapy of dental caries. This  
CC sequence represents a Streptococcus mutans monoclonal antibody-related  
CC protein  
XX  
XX Sequence 1375 AA;  
SQ

Query Match 100.0%; Score 108; DB 5; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 7.9e-08;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVADLLQI 22  
Db 468 DANFDSIRVDVNDVADLLQI 489  
|||||

RESULT 9  
ADD93655  
ID ADD93655 standard; protein; 1375 AA.  
XX  
XX ADD93655;  
XX  
XX 29-JAN-2004 (first entry)  
XX  
XX Streptococcus mutans glucosyltransferase-C.  
XX  
XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.  
XX  
XX Streptococcus mutans.  
XX  
XX WO2003075845-A2.  
XX  
XX 18-SEP-2003.  
XX  
XX 07-MAR-2003; 2003WO-US006962.  
XX  
XX 07-MAR-2002; 2002US-0363209P.  
XX  
XX 08-AUG-2002; 2002US-0402483P.  
XX  
XX (FORS-) FORSYTH INST.  
XX  
XX Smith DJ, Taubman MA;  
XX  
XX WPI; 2003-845091/78.  
XX  
XX Composition useful as vaccines for dental caries comprises a fragment of  
PT a glucan binding protein-B binding to a major histocompatibility complex  
PT class II protein.  
XX  
XX Claim 16; Page 13; 49pp; English.  
XX  
XX The present sequence is the protein sequence of Streptococcus mutans

CC glucosyltransferase-C (GTF-C). Peptide fragments of GTF-C, especially  
CC from the catalytic domain of the polypeptide, can be used in immunogenic  
CC compositions and subunit vaccines for dental caries. These compositions  
CC comprise a major histocompatibility complex (MHC) class II protein-  
CC binding peptide from S. mutans glucan binding protein-B (GbpB)  
CC covalently linked with a peptide fragment of a streptococcal  
CC glucosyltransferase. The compositions are used in a claimed method of  
CC eliciting production of an antibody in a mammal. Diepitopic or  
CC multi-epitopic polypeptides can be prepared synthetically or by  
CC recombinant DNA technology. Antibodies raised against MHC class II  
CC binding fragments of GbpB can be used in passive immunisation.  
XX  
XX Sequence 1375 AA;  
SQ

Query Match 100.0%; Score 108; DB 7; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 7.9e-08;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVADLLQI 22  
Db 468 DANFDSIRVDVNDVADLLQI 489  
|||||

RESULT 10  
ADX37278  
ID ADX37278 standard; protein; 1375 AA.  
XX  
XX ADX37278;  
XX  
XX 21-APR-2005 (first entry)  
XX  
XX Streptococcus mutant glucan binding protein B variant #7.  
XX  
XX immunogenicity; immune stimulation; glucan binding protein-B;  
KW microparticle; major histocompatibility complex; tooth disease.  
XX  
XX Streptococcus mutans.  
XX  
XX US2005031633-A1.  
XX  
XX 10-FEB-2005.  
XX  
XX 09-MAR-2004; 2004US-00797821.  
XX  
XX 13-APR-1998; 98US-0081550P.  
PR 08-JAN-1999; 99US-0115142P.  
PR 12-APR-1999; 99US-00290049.  
PR 07-MAR-2002; 2002US-0363209P.  
PR 08-AUG-2002; 2002US-0402483P.  
PR 07-MAR-2003; 2003US-00383930.  
XX  
XX (SMIT/) SMITH D J.  
PA (TAUB/) TAUBMAN M A.  
XX  
XX Smith DJ, Taubman MA;  
XX  
XX WPI; 2005-151644/16.  
XX  
XX New composition comprising a fragment of a glucan binding protein-B  
PT (GbpB) that binds to MHC class II protein, and a biocompatible  
PT microparticle, useful for producing an antibody (claimed) for immunizing  
PT mammals against dental caries.  
XX  
XX Claim 7; SEQ ID NO 35; 73pp; English.  
XX  
XX The invention relates to a composition comprising a fragment of a glucan  
CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
CC fragment binds to a major histocompatibility complex (MHC) class II  
CC protein. The composition is useful for producing an antibody for  
CC immunizing mammals against dental caries. This sequence corresponds to a  
CC Streptococcus mutans GbpB protein of the invention.  
XX  
XX Sequence 1375 AA;  
SQ

```

Query Match      100.0%; Score 108; DB 9; Length 1375;
Best Local Similarity 100.0%; Pred. No. 7.9e-08;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22
    |||||
Db 468 DANFDSIRVDVNDVADLLQI 489

RESULT 11
AAU98036
ID AAU98036 standard; protein; 1475 AA.
XX
AC AAU98036;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFB mutant D567T/D571K.
XX
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
XX
US2002031826-A1.
XX
14-MAR-2002.
XX
19-DEC-2000; 2000US-00740274.
XX
07-JUN-1995; 95US-00478704.
XX
07-JUN-1995; 95US-00482711.
XX
07-JUN-1995; 95US-00485243.
XX
16-JAN-1998; 98US-00007999.
XX
16-JAN-1998; 98US-00008172.
XX
20-JAN-1998; 98US-00009620.
XX
11-DEC-1998; 98US-00210361.
XX
(NICH/) NICHOLS S E.
XX
PI Nichols SE;
XX
WPI; 2002-414332/44.
XX
Glucosyltransferase B or D protein useful for producing a glucan useful
as substitutes for and additions to modified starch and latexes in paper
manufacture, comprises mutations in specific positions.
XX
Claim 36; Page; 44pp; English.
XX
The invention an isolated protein comprising a glucosyltransferase (GTF)
B polypeptide having changes at position from I448V, D457N, D567T,
K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a
GTF D polypeptide having changes at positions from T589D, T589E, N471D,
N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
complementary polynucleotide, a ribonucleic acid sequence encoding the
GTF mutant, an expression cassette comprising the polynucleotide operably
linked to a promoter, a vector comprising the expression cassette, host
cell introduced with the vector, a transgenic plant comprising the
vector, a seed or tuber from the transgenic plant, a paper sizing and/or
coating composition comprising a glucan produced in a plant transformed
with a gene encoding the mutant GTF, wild type or, starch, a latex,

```

```

thermoplastic molecule or their combinations or glucan and starch where
the glucan is produced in the amyloplast and/or vacuole or a maize line
deficient in starch biosynthesis, transformed with a gene encoding a
glucosyltransferase B or D enzyme, wild-type or mutant and a paper
comprising the glucan (paper sizing/coating agent). The vector is useful
for producing a glucan in a plant. The method comprises transforming a
plant cell with the vector, growing the plant cell under plant growing
conditions to produce a regenerated plant and inducing expression of the
polynucleotide for a time sufficient to produce the glucan in the
regenerated plant, where the vector contains a transit sequence from
ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
chlorophyll AB binding protein to produce a transgenic plant, and glucan
is produced in the amyloplast of potato or the vacuole of sugar beet.
Glucans are useful as substitutes for and additions to modified starch
and latexes in paper manufacture. Unlike prior art techniques, which
require input materials that produce chemical effluents, paper
manufacture utilising the glucan produced by GTF, which utilises
biologically produced input materials, is more cost-effective and
environmentally friendly. Moreover, glucans also exhibit thermoplastic
properties and impart gloss to the paper during coating step. The present
sequence represents a GTFB mutant of the invention. Note: The present
sequence is not shown in the specification but was created by the indexer
using the GTFB sequence appearing as AAU98027 and the information in
claim 36
XX
SQ Sequence 1475 AA;

```

```

Query Match      100.0%; Score 108; DB 5; Length 1475;
Best Local Similarity 100.0%; Pred. No. 8.6e-08;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 DANFDSIRVDVNDVADLLQI 22
    |||||
Db 442 DANFDSIRVDVNDVADLLQI 463

```

```

RESULT 12
AAU98037

```

```

ID AAU98037 standard; protein; 1475 AA.

```

```

XX
AC AAU98037;

```

```

DT 27-AUG-2002 (first entry)

```

```

DE S. mutans glucosyltransferase GTFB mutant D567T/D571K/K1014T.

```

```

KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.

```

```

XX Streptococcus mutans.

```

```

OS Synthetic.

```

```

XX Key Location/Qualifiers

```

```

FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"

```

```

FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"

```

```

FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"

```

```

XX US2002031826-A1.

```

```

PN 14-MAR-2002.

```

```

XX 19-DEC-2000; 2000US-00740274.

```

```

XX 07-JUN-1995; 95US-00478704.

```

```

XX 07-JUN-1995; 95US-00482711.

```

```

XX 07-JUN-1995; 95US-00485243.

```

```

XX 16-JAN-1998; 98US-00007999.

```

```

XX 16-JAN-1998; 98US-00008172.

```

```

XX 20-JAN-1998; 98US-00009620.

```



PR 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 PA  
 XX Nichols SE;  
 XX WPI; 2002-414332/44.  
 DR  
 XX  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 XX Claim 36; Page; 44pp; English.  
 PS  
 XX  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D, and  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX  
 SQ Sequence 1475 AA;  
 Query Match 100.0%; Score 108; DB 5; Length 1475;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-08;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DANFDSIRVDANVDADLLQI 22  
 Db 442 DANFDSIRVDANVDADLLQI 463  
 RESULT 13  
 AAU98040  
 ID AAU98040 standard; protein; 1475 AA.  
 XX  
 AC AAU98040;  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE S. mutans glucosyltransferase GTFB mutant K779Q.

XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture; mutant; muten.  
 XX  
 OS Streptococcus mutans.  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FH Misc-difference 779  
 FT /note= "Wild-type Lys substituted by Gln"  
 XX  
 XX US2002031826-A1.  
 XX  
 XX 14-MAR-2002.  
 XX  
 XX 19-DEC-2000; 2000US-00740274.  
 XX  
 XX 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 XX (NICH/) NICHOLS S E.  
 PA  
 XX Nichols SE;  
 XX  
 XX WPI; 2002-414332/44.  
 DR  
 XX  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 XX Claim 36; Page; 44pp; English.  
 PS  
 XX  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D, and  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX  
 SQ Sequence 1475 AA;  
 Query Match 100.0%; Score 108; DB 5; Length 1475;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-08;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DANFDSIRVDANVDADLLQI 22  
 Db 442 DANFDSIRVDANVDADLLQI 463  
 RESULT 13  
 AAU98040  
 ID AAU98040 standard; protein; 1475 AA.  
 XX  
 AC AAU98040;  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE S. mutans glucosyltransferase GTFB mutant K779Q.

CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36

XX SQ Sequence 1475 AA;

Query Match 100.0%; Score 108; DB 5; Length 1475;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-08;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DANFDSIRVDVNDVADLLQI 22  
 |||||  
 Db 442 DANFDSIRVDVNDVADLLQI 463

RESULT 14

AAU98033  
 ID AAU98033 standard; protein; 1475 AA.

AC AAU98033;

XX 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFB mutant K1014T.  
 XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX Streptococcus mutans.

OS Synthetic.

XX Key Location/Qualifiers

XX Misc-difference 1014 /note= "wild-type lys substituted by Thr"

FT US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

PI WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful

PT as substitutes for and additions to modified starch and latexes in paper

PT manufacture, comprises mutations in specific positions.

XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)

CC B polypeptide having changes at position from I448V, D457N, D567T,

CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,

CC I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a

CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,

CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the

CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its

CC complementary polynucleotide, a ribonucleic acid sequence encoding the

CC GTF mutant, an expression cassette comprising the polynucleotide operably

CC linked to a promoter, a vector comprising the expression cassette, host

CC cell introduced with the vector, a transgenic plant comprising the

CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, paper  
 CC require input materials that produce chemical effluents, which  
 CC manufacture utilizing the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36

XX Sequence 1475 AA;

Query Match 100.0%; Score 108; DB 5; Length 1475;  
 Best Local Similarity 100.0%; Pred. No. 8.6e-08;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DANFDSIRVDVNDVADLLQI 22

Db 442 DANFDSIRVDVNDVADLLQI 463

RESULT 15

AAU98039  
 ID AAU98039 standard; protein; 1475 AA.

AC AAU98039;

XX 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFB mutant YYY169-171AAA.

XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;

KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX Streptococcus mutans.

OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 169..171

FT /note= "Wild-type Tyr-Tyr substituted by Ala-Ala-Ala"

XX US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.  
XX PI Nichols SE;  
XX WPI; 2002-414332/44.  
XX

XX Glucosyltransferase B or D protein useful for producing a glucan useful  
PT as substitutes for and additions to modified starch and latexes in paper  
PT manufacture, comprises mutations in specific positions.  
XX

PS Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
CC B polypeptide having changes at position from I448V, D457N, D567T,  
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
CC I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
CC GTF mutant, an expression cassette comprising the polynucleotide operably  
CC linked to a promoter, a vector comprising the expression cassette, host  
CC cell introduced with the vector, a transgenic plant comprising the  
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
CC coating composition comprising a glucan produced in a plant transformed  
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
CC thermoplastic molecule or their combinations or glucan and starch where  
CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
CC deficient in starch biosynthesis, transformed with a gene encoding a  
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
CC comprising the glucan (paper sizing/coating agent). The vector is useful  
CC for producing a glucan in a plant. The method comprises transforming a  
CC plant cell with the vector, growing the plant cell under plant growing  
CC conditions to produce a regenerated plant and inducing expression of the  
CC polynucleotide for a time sufficient to produce the glucan in the  
CC regenerated plant, where the vector contains a transit sequence from  
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
CC Glucans are useful as substitutes for and additions to modified starch  
CC and latexes in paper manufacture. Unlike prior art techniques, which  
CC require input materials that produce chemical effluents, paper  
CC manufacture utilising the glucan produced by GTF, which utilises  
CC biologically produced input materials, is more cost-effective and  
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
CC properties and impart gloss to the paper during coating step. The present  
CC sequence represents a GTFB mutant of the invention. Note: The present  
CC sequence is not shown in the specification but was created by the indexer  
CC using the GTFB sequence appearing as AAU98027 and the information in  
CC claim 36  
XX

SQ Sequence 1475 AA;

Query Match 100.0%; Score 108; DB 5; Length 1475;  
Best Local Similarity 100.0%; Pred. NO. 8.6e-08;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22  
DB 442 DANFDSIRVDVNDVADLLQI 463

Search completed: February 10, 2006, 22:19:42  
Job time : 15.9234 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:20:08 ; Search time 2.59476 Seconds  
(without alignments)  
815.787 Million cell updates/sec

Title: US-10-797-821-25  
Perfect score: 108  
Sequence: 1 DANFDSIRVDVNDVDADLLQI 22  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues  
Total number of hits satisfying chosen parameters: 283416  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : PIR 80.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	100.0	1375	2 J0345	dextranucrase (EC
2	108	100.0	1475	2 B3135	gtfB protein precu
3	100	92.6	1518	2 A4481	glucosyltransferas
4	98	90.7	1365	2 A41483	glucosyltransferas
5	98	90.7	1508	2 T31098	probable dextranu
6	96	88.9	1577	2 T30858	glucosyltransferas
7	96	88.9	1592	2 A38175	glucosyltransferas
8	94	87.0	1431	2 A45866	dextranucrase (EC
9	93	86.1	1290	2 JC5473	dextranucrase (EC
10	93	86.1	1449	2 T30857	glucosyltransferas
11	93	86.1	1449	2 T30552	glucosyltransferas
12	85	78.7	1599	2 S22737	glucosyltransferas
13	54	50.0	655	1 ALKEG	cyclomaltodextrin
14	46.5	43.1	762	2 C69657	cobalamin-independ
15	46	42.6	128	2 AC1607	transcription term
16	46	42.6	673	2 B70528	probable peptidase
17	46	42.6	826	2 T06494	1,4-alpha-glucan b
18	45.5	42.1	147	2 D69732	PBSX prophage ORF
19	45.5	42.1	378	1 A40004	histidine decarbox
20	45.5	42.1	512	2 S61905	threonine synthase
21	45	41.7	227	2 H69453	hypothetical prote
22	45	41.7	457	2 G82925	hypothetical prote
23	44	40.7	280	2 E72296	pantoate-beta-alan
24	44	40.7	315	2 AF1393	glycosyl transfera
25	44	40.7	315	2 AF1768	glycosyl transfera
26	44	40.7	594	2 S37816	hypothetical prote
27	44	40.7	1819	2 D97033	uncharacterized pr
28	43.5	40.3	307	2 D84536	hypothetical prote
29	43.5	40.3	550	2 A82378	conserved hypothet

30 43 39.8 9 2 A39841 sucrose 3-glucoseyl  
31 43 39.8 128 2 AG1444 transcription term  
32 43 39.8 352 2 G95895 probable methyla  
33 43 39.8 385 1 S48456 histidinol-phospha  
34 43 39.8 448 2 B72262 glucose-6-phosphat  
35 43 39.8 451 2 F97230 probable maltodext  
36 43 39.8 527 2 F84256 IMP dehydrogenase  
37 43 39.8 638 2 S41159 sodium transport p  
38 43 39.8 668 2 S49639 probable membrane  
39 43 39.8 756 2 F83704 homocystein methyl  
40 42.5 39.4 147 2 H69947 phase-related prot  
41 42.5 39.4 486 2 AG0355 probable membrane  
42 42.5 39.4 1214 2 H75034 reverse gyrase (to  
43 42.5 39.4 1624 2 C71129 probable reverse g  
44 42 38.9 179 2 S35942 probable ATP synth  
45 42 38.9 179 2 S48643 ATP synthase - soy

ALIGNMENTS

RESULT 1  
J0345  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)  
N;Alternate names: sucrose 6-glucosyltransferase  
C;Species: Streptococcus mutans  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
C;Accession: J0345; C3135  
R;Ueda, S.; Shiroza, T.; Kuramitsu, H.K.  
Gene 69, 101-109, 1988  
A;Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.  
A;Reference number: J0345; MUID:89137980; PMID:2976010  
A;Accession: J0345  
A;Molecule type: DNA  
A;Residues: 1-1375 <UED>  
A;Cross-references: UNIPROT:P13470; UNIPARC:UPI0000155515  
A;Experimental source: GS-5  
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A;Reference number: A3135; MUID:87308013; PMID:3040685  
A;Accession: C3135  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-349 <SHI>  
A;Cross-references: UNIPARC:UPI000014E25D; GB:M17361  
C;Genetics:  
A;Gene: gtfC  
C;Function:  
A;Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
C;Keywords: duplication; glycosyltransferase; hexosyltransferase  
F;1-34/Domain: signal sequence #status predicted <SIG>  
F;35-1375/Product: glucosyltransferase #status predicted <MAT>  
F;1126-1145/Domain: cpl repeat homology <CP1>  
F;1253-1272/Domain: cpl repeat homology <CP2>  
F;1318-1337/Domain: cpl repeat homology <CP3>

Query Match 100.0%; Score 108; DB 2; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 6.9e-08;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVDADLLQI 22  
|||||  
Db 468 DANFDSIRVDVNDVDADLLQI 489

RESULT 2  
B3135  
gtfB protein precursor - Streptococcus mutans  
C;Species: Streptococcus mutans  
C;Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 09-Jul-2004  
C;Accession: B3135; A33128  
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987  
A;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A;Reference number: A33135; MUID: 97308013; PMID: 3040685  
A;Accession: B33135  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1475 <SHI>  
A;Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:g153639; PIDN:R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
submitted to the Protein Sequence Database, September 1990  
A;Reference number: A33128  
A;Accession: A33128  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-171,173-641,'N' 643-1475 <SH2>  
A;Cross-references: UNIPARC:UPI000017ACSE  
A;Experimental source: strain GS-5  
F:1096-1115/Domain: cpl repeat homology <CP1>  
F:1224-1243/Domain: cpl repeat homology <CP2>  
F:1289-1308/Domain: cpl repeat homology <CP3>  
F:1354-1373/Domain: cpl repeat homology <CP4>  
F:1419-1438/Domain: cpl repeat homology <CP5>  
  
Query Match 100.0%; Score 108; DB 2; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 7.5e-08;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 DANFDSIRVDVNDVADLLQI 22  
Db 442 DANFDSIRVDVNDVADLLQI 463  
  
RESULT 3  
A44811  
Glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius  
C;Species: Streptococcus salivarius  
C;Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 09-Jul-2004  
C;Accession: A44811; S22726; S28809  
R;Giffard, P.M.; Simpson, C.L.; Millward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A;Title: Molecular characterization of a cluster of at least two glucosyltransferase genes  
A;Reference number: A44811; MUID: 92148377; PMID: 1838391  
A;Accession: A44811  
A;Molecule type: DNA  
A;Residues: 1-1518 <GIF>  
A;Cross-references: UNIPROT:Q00600; UNIPARC:UPI00000BEF31; EMBL:Z11873; NID:g47526; PIDN:A;Note: Sequence extracted from NCBI backbone (NCBIN:81050, NCBI:P:81052)  
C;Genetics:  
A;Gene: gtfJ  
C;Keywords: glycosyltransferase; hexosyltransferase  
F:1307-1326/Domain: cpl repeat homology <CP4>  
  
Query Match 92.6%; Score 100; DB 2; Length 1518;  
Best Local Similarity 86.4%; Pred. No. 1.2e-06;  
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 DANFDSIRVDVNDVADLLQI 22  
Db 462 DANFDSIRVDVNDVADMLQI 483  
  
RESULT 4  
A41483  
Glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus  
C;Species: Streptococcus sobrinus  
C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Mar-2004  
C;Accession: A41483  
R;Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
Infect. Immun. 58, 2452-2458, 1990  
A;Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltransferase  
A;Reference number: A41483; MUID: 90316665; PMID: 2142479  
A;Accession: A41483  
A;Molecule type: DNA

A;Residues: 1-1365 <GIL>  
A;Cross-references: UNIPARC:UPI000012BCB6; GB:M30943; NID:g153652; PIDN:AAA26898.1; PIDN:R;C;Genetics:  
A;Gene: gtfS  
C;Keywords: glycosyltransferase; hexosyltransferase  
  
Query Match 90.7%; Score 98; DB 2; Length 1365;  
Best Local Similarity 86.4%; Pred. No. 2.1e-06;  
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 DANFDSIRVDVNDVADLLQI 22  
Db 428 DANFDGVRVDVNDVNDLLQI 449  
  
RESULT 5  
T31098  
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides  
C;Species: Leuconostoc mesenteroides  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T31098  
R;Monchois, V.; Remaud-Simeon, M.; Monsan, P.; Willemot, R.M.  
FEMS Microbiol. Lett. 159, 307-315, 1998  
A;Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (DSF)  
A;Reference number: Z20981; MUID: 98164374; PMID: 9503626  
A;Accession: T31098  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1508 <MON>  
A;Cross-references: UNIPROT:O52224; UNIPARC:UPI00000BB69B; EMBL:AF030129; NID:g2766611; PIDN:A;Experimental source: strain NRRL B-1299  
C;Genetics:  
A;Gene: dsrB  
C;Function:  
A;Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
C;Keywords: glycosyltransferase; hexosyltransferase  
  
Query Match 90.7%; Score 98; DB 2; Length 1508;  
Best Local Similarity 95.2%; Pred. No. 2.4e-06;  
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 ANFDSIRVDVNDVADLLQI 22  
Db 525 ANFDGIRVDVNDVADLLQI 545  
  
RESULT 6  
T30858  
glucosyltransferase - Streptococcus salivarius  
C;Species: Streptococcus salivarius  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T30858  
R;Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A;Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for proteins  
A;Reference number: Z20909; MUID: 95122197; PMID: 7822030  
A;Accession: T30858  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1577 <SIM>  
A;Cross-references: UNIPROT:Q55265; UNIPARC:UPI00000B8087; EMBL:L35928; NID:g662380; PIDN:R;C;Genetics:  
A;Gene: gtfm  
  
Query Match 88.9%; Score 96; DB 2; Length 1577;  
Best Local Similarity 85.7%; Pred. No. 5e-06;  
Matches 18; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
QY 2 ANFDSIRVDVNDVADLLQI 22  
Db 550 ANFDGVRVDVNDVNDLLQI 570

```
RESULT 7
glucosyltransferase precursor - Streptococcus sobrinus
C;Species: Streptococcus sobrinus
C;Date: 28-Aug-1992 #sequence_revision 28-Aug-1992 #text_change 15-Mar-2004
C;Accession: A38175
R;Abou, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.
J. Bacteriol. 173, 989-996, 1991
A;Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus
A;Reference number: A38175; MUID:91123227; PMID:1704006
A;Accession: A38175
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1592 <ABO>
A;Cross-references: UNIPARC:UPI000012BCB2; GB:D90213; NID:G217032; PIDN:BAA14241.1; PID:
F;1093-1112/Domain: cpl repeat homology <CP1>
F;1222-1241/Domain: cpl repeat homology <CP2>
F;1287-1306/Domain: cpl repeat homology <CP3>
F;1330-1351/Domain: cpl repeat homology <CP4>
F;1352-1371/Domain: cpl repeat homology <CP5>
F;1402-1420/Domain: cpl repeat homology <CP6>
F;1465-1484/Domain: cpl repeat homology <CP7>
F;1513-1532/Domain: cpl repeat homology <CP8>

Query Match      88.9%; Score 96; DB 2; Length 1592;
Best Local Similarity 90.9%; Pred. No. 5.1e-06;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQI 22
    |||||
Db 438 DANFDSIRVDVNDVNDADLLQI 459

RESULT 8
dextranase (EC 2.4.1.5) precursor - Streptococcus mutans
C;Species: Streptococcus mutans
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C;Accession: A45866
R;Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A;Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl-
A;Reference number: A45866; MUID:9110958; PMID:2148600
A;Accession: A45866
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1431 <HON>
A;Cross-references: UNIPARC:UPI000017AC5C; GB:M29296
C;Keywords: glycosyltransferase; hexosyltransferase
F;181-201/Domain: cpl repeat homology <CP1>
F;1127-1146/Domain: cpl repeat homology <CP2>
F;1192-1211/Domain: cpl repeat homology <CP3>
F;1257-1276/Domain: cpl repeat homology <CP4>
F;1277-1297/Domain: cpl repeat homology <CP5>
F;1321-1340/Domain: cpl repeat homology <CP6>
F;1341-1361/Domain: cpl repeat homology <CP6>
F;1385-1404/Domain: cpl repeat homology <CP7>

Query Match      87.0%; Score 94; DB 2; Length 1431;
Best Local Similarity 81.8%; Pred. No. 9e-06;
Matches 18; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQI 22
    :|||:|||||:|||||
Db 456 EANFDGVRVDVNDVNDADLLQI 477

RESULT 9
JC5473
dextranase (EC 2.4.1.5) - Leuconostoc mesenteroides
C;Species: Leuconostoc mesenteroides
C;Date: 07-Jul-1997 #sequence_revision 29-Aug-1997 #text_change 09-Jul-2004
C;Accession: JC5473
```

```
R;Monchois, V.; Willemot, R.M.; Renaud-Simeon, M.; Croux, C.; Monsan, P.
Gene 182, 23-32, 1996
A;Title: Cloning and sequencing of a gene coding for a novel dextranase from Leucono
A;Reference number: JC5473; MUID:97113686; PMID:8982063
A;Accession: JC5473
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1290 <NON>
A;Cross-references: UNIPROT:Q48756; UNIPARC:UPI000017ABDF; GB:U38181
C;Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose ont
C;Genetics:
A;Gene: dsrA
C;Keywords: glycosyltransferase; hexosyltransferase
F;78-870/Domain: catalytic #status predicted <CAT>
F;922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match      86.1%; Score 93; DB 2; Length 1290;
Best Local Similarity 90.5%; Pred. No. 1.1e-05;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ANFDSIRVDVNDVNDADLLQI 22
    |||||
Db 278 ANFDGYRVDVNDVNDADLLQI 298

RESULT 10
T30857
glucosyltransferase - Streptococcus salivarius
C;Species: Streptococcus salivarius
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C;Accession: T30857
R;Simpson, C.L.; Giffard, P.M.; Jacques, N.A.
Infect. Immun. 63, 609-621, 1995
A;Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri
A;Reference number: Z20909; MUID:95122197; PMID:7822030
A;Accession: T30857
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1449 <SIM>
A;Cross-references: UNIPROT:Q55264; UNIPARC:UPI00000B166E; EMBL:L35495; NID:G662378; PID
C;Genetics:
A;Gene: gtfL

Query Match      86.1%; Score 93; DB 2; Length 1449;
Best Local Similarity 81.8%; Pred. No. 1.3e-05;
Matches 18; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQI 22
    |||||
Db 497 DENFDGVRVDVNDVNDADLLQI 518

RESULT 11
T30552
glucosyltransferase N - Streptococcus salivarius (fragment)
C;Species: Streptococcus salivarius
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
C;Accession: T30552
R;Jaffe, R.I.
Submitted to the EMBL Data Library, February 1998
A;Description: Streptococcus salivarius V1477 gtfN.
A;Reference number: Z20854
A;Accession: T30552
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-1449 <JAF>
A;Cross-references: UNIPROT:O68542; UNIPARC:UPI00000B10FD; EMBL:AF049609; NID:G2935545;
C;Genetics:
A;Gene: gtfN

Query Match      86.1%; Score 93; DB 2; Length 1449;
Best Local Similarity 81.8%; Pred. No. 1.3e-05;
Matches 18; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

QY 1 DANFDSIRVDVNDVADLLQI 22  
Db 497 DENFDGVRVDVNDVADLLQI 518

RESULT 12  
S22737  
glucosyltransferase (EC 2.4.1.1-) S - Streptococcus salivarius  
C;Species: Streptococcus salivarius  
C;Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 09-Jul-2004  
C;Accession: S22737; S28810; B44811; S22727  
R;Jacques, N.  
submitted to the EMBL Data Library, March 1992  
A;Reference number: S22726  
A;Accession: S22737  
A;Molecule type: DNA  
A;Residues: 1-1599 <JAC>  
A;Cross-references: UNIPROT:Q00599; UNIPARC:UPI00000BFF34; EMBL:Z11872; NID:g47530; PIDN  
A;Experimental source: ATCC 25975  
R;Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A;Title: Molecular characterization of a cluster of at least two glucosyltransferase gen  
A;Reference number: A44811; MUID:92148377; PMID:1838391  
A;Accession: S28810  
A;Molecule type: DNA  
A;Residues: 1-51 <GIF>  
A;Cross-references: UNIPARC:UPI000017027C; EMBL:Z11873  
C;Genetics:  
C;Gene: gtfK  
C;Keywords: glycosyltransferase; hexosyltransferase  
P;1456-1475/Domain: cpl repeat homology <CPR>

Query Match 78.7%; Score 85; DB 2; Length 1599;  
Best Local Similarity 72.7%; Pred. No. 0.00023;  
Matches 16; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22  
Db 452 DAHFDGIRVDVNDVADLLQI 473

RESULT 13  
ALXBG  
cyclomaltodextrin glucanotransferase (EC 2.4.1.19) precursor - Klebsiella pneumoniae  
C;Species: Klebsiella pneumoniae  
C;Date: 31-Mar-1993 #sequence\_revision 31-Mar-1993 #text\_change 31-Dec-2004  
C;Accession: A29023; B29023  
R;Binder, F.; Huber, O.; Bock, A.  
Gene 47, 269-277, 1986  
A;Title: Cyclodextrin-glycosyltransferase from Klebsiella pneumoniae M5a1: cloning, nucl  
A;Reference number: A29023; MUID:87163498; PMID:2951300  
A;Accession: A29023  
A;Molecule type: DNA  
A;Residues: 1-655 <BIN>  
A;Cross-references: UNIPROT:P08704; UNIPARC:UPI0000127416; EMBL:M15264; NID:g149178; PID  
A;Note: the authors translated the codon CAG for residue 233 as Asn  
A;Accession: B29023  
A;Molecule type: protein  
A;Residues: 31-33 <BI2>  
A;Cross-references: UNIPARC:UPI0000172935  
C;Genetics:  
A;Gene: cgt  
C;Function:  
A;Description: catalyzes the breaking and reformation of 1,4-alpha-D-glucopyranosyl bond  
A;Note: can also catalyze the disproportionation transfer of 6 to 8 1,4-alpha-D-glucopyr  
C;Superfamily: cyclomaltodextrin glucanotransferase/alpha-amylase; alpha-amylase core h  
C;Keywords: calcium; duplication; extracellular protein; glycosyltransferase; hexosyltra  
P;1-30/Domain: signal sequence #status predicted<SIG>  
F;31-655/Product: cyclomaltodextrin glucanotransferase #status experimental <MAT>  
F;33-163/Domain: A1 <DA1>  
F;164-226/Domain: B <DOB>  
F;221-366/Domain: alpha-amylase core homology <AMY>

F;227-416/Domain: A2 <DA2>  
F;417-555/Domain: C and D <DOC>  
F;556-655/Domain: E <DOE>  
F;55,57,60,61,79,81/Binding site: calcium (Asp, Asn, Gly, Asp) #status predicted  
F;164,214,223,257/Binding site: calcium (Asn, Lys, Asp, His) #status predicted  
F;253,287,363/Active site: Asp, Glu, Asp #status predicted

Query Match 50.0%; Score 54; DB 1; Length 655;  
Best Local Similarity 42.9%; Pred. No. 3.5;  
Matches 9; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 21  
Db 244 DAGVDAIRDAIKHMDKSFQI 264

RESULT 14  
C69657  
cobalamin-independent methionine synthase metC - Bacillus subtilis  
C;Species: Bacillus subtilis  
C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C;Accession: C69657  
R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997  
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler  
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.;  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,  
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle,  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon,  
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, P.; Sekiguchi, J.; Sekowska, A.; Seror,  
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.  
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
A;Reference number: A69580; MUID:98044033; PMID:9384377  
A;Accession: C69657  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-762 <KUN>  
A;Cross-references: UNIPROT:P80877; UNIPARC:UPI00000602B5; GB:Z99110; GB:AL009126; NID:g  
A;Experimental source: strain 168  
C;Genetics:  
C;Superfamily: cobalamin-independent methionine synthase

Query Match 43.1%; Score 46.5; DB 2; Length 762;  
Best Local Similarity 42.9%; Pred. No. 56;  
Matches 9; Conservative 8; Mismatches 3; Indels 1; Gaps 1;

QY 2 ANFDSIRVDVNDVADLLQI 22  
Db 649 SNFEDI-VDTINDLDADVITI 668

RESULT 15  
AC1607  
transcription termination protein (NusB) homolog lin1396 [imported] - Listeria innocua (E  
C;Species: Listeria innocua  
C;Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
C;Accession: AC1607  
R;Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,  
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl, H.,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A;Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Mat  
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,  
A;Title: Comparative genomics of Listeria species.  
A;Reference number: AB1077; MUID:21537279; PMID:11679669  
A;Accession: AC1607  
A;Status: preliminary



A;Molecule type: DNA  
A;Residues: 1-128 <GLA>  
A;Cross-references: UNIPROT:Q92B25; UNIPARC:UPI0000130A5B; GB:AL592022; PIDN:CAC96627.1  
A;Experimental source: strain Clip11262  
C;Genetics:  
A;Gene: lin1396  
C;Superfamily: nusB protein  
C;Keywords: transcription termination

Query Match 42.6%; Score 46; DB 2; Length 128;  
Best Local Similarity 36.4%; Pred. No. 8.6;  
Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22  
Db 59 EPNLDNWRIDRLNKVDLSLRL 80

Search completed: February 10, 2006, 22:33:14  
Job time : 2.71976 secs

us-10-797-821-25.rpr

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:07:53 ; Search time 15.1028 Seconds  
(without alignments)  
1027.730 Million cell updates/sec

Title: US-10-797-821-25  
Perfect score: 108  
Sequence: 1 DANFDSIRVDVNDVDALLQI 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 segs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108	100.0	1016	2 Q9LCJ7 LEUME	Q9LCJ7 leuconostoc
2	108	100.0	1454	2 Q69A94 LEUME	Q69A94 leuconostoc
3	108	100.0	1455	1 GTFC STRMU	P13470 streptococc
4	108	100.0	1476	1 GTFB STRMU	P08987 streptococc
5	108	100.0	1590	2 Q55263 9STRE	Q55263 streptococc
6	108	100.0	1590	2 Q59983 9STRE	Q59983 streptococc
7	108	100.0	1597	1 GTFL STRDO	F11001 streptococc
8	104	96.3	1522	2 Q6TXV4 LEUME	Q6TXV4 leuconostoc
9	104	96.3	1527	2 Q8KRE1 LEUME	Q8KRE1 leuconostoc
10	104	96.3	1527	2 Q9ZAR4 LEUME	Q9ZAR4 leuconostoc
11	103	95.4	1772	2 Q5SBN0 LACRE	Q5SBN0 lactobacill
12	102	94.4	1463	2 Q5SBN6 LACFE	Q5SBN6 lactobacill
13	100	92.6	1518	2 Q00600 STRSL	Q00600 streptococc
14	98	90.7	1330	2 Q84CN4 LEUME	Q84CN4 leuconostoc
15	98	90.7	1365	1 GTFS STRDO	P29336 streptococc
16	98	90.7	1477	2 Q9L466 LEUME	Q9L466 leuconostoc
17	98	90.7	1508	2 Q9EZH5 LEUME	Q9EZH5 leuconostoc
18	98	90.7	1508	2 Q52224 LEUME	Q52224 leuconostoc
19	97	89.8	1512	2 Q9WXJ5 9STRE	Q9WXJ5 streptococc
20	97	89.8	1554	2 Q8KZL5 9STRE	Q8KZL5 streptococc
21	96	88.9	1577	2 Q55265 STRSL	Q55265 streptococc
22	96	88.9	1592	1 GTF2 STRDO	P27470 streptococc
23	96	88.9	1595	2 Q5SBN3 LACSK	Q5SBN3 lactobacill
24	95	88.0	2057	2 Q9RE05 LEUME	Q9RE05 leuconostoc
25	94	87.0	1462	1 GTFD STRMU	P49331 streptococc
26	93	86.1	1290	2 Q48756 LEUME	Q48756 leuconostoc
27	93	86.1	1338	2 Q9WXJ4 9STRE	Q9WXJ4 streptococc
28	93	86.1	1449	2 Q68542 STRSL	Q68542 streptococc
29	93	86.1	1449	2 Q55264 STRSL	Q55264 streptococc
30	93	86.1	2835	2 Q8G9Q2 LEUME	Q8G9Q2 leuconostoc
31	92	85.2	1575	2 Q9LCH3 STROR	Q9LCH3 streptococc

32	92	85.2	1577	2 Q54178 STRGN	Q54178 streptococc
33	91	84.3	1772	2 Q5SBN3 LACRE	Q5SBN3 lactobacill
34	88	81.5	1561	2 Q5SBN8 9LACO	Q5SBN8 lactobacill
35	87	80.6	1506	2 Q56CX8 9STRE	Q56CX8 streptococc
36	87	80.6	1781	2 Q5SBL9 LACRE	Q5SBL9 lactobacill
37	87	80.6	1781	2 Q4JCS4 LACRE	Q4JCS4 lactobacill
38	87	80.6	1781	2 Q4JLC7 LACRE	Q4JLC7 lactobacill
39	85	78.7	1599	2 Q00599 STRSL	Q00599 streptococc
40	79	73.1	1231	2 Q5SBN1 LACRE	Q5SBN1 lactobacill
41	77	71.3	1619	2 Q5SBN0 LACRE	Q5SBN0 lactobacill
42	65	60.2	522	2 Q8VV10 STRSA	Q8VV10 streptococc
43	54	50.0	655	1 CDGT KLEOX	P08704 klebsiella
44	51.5	47.7	762	2 Q6SKT8 BACLD	Q6SKT8 bacillus li
45	50	46.3	4405	2 Q8ILZ2 FLAF7	Q8ILZ2 plasmodium

ALIGNMENTS

RESULT 1

Q9LCJ7 LEUME  
ID Q9LCJ7 LEUME PRELIMINARY; PRT; 1016 AA.  
AC Q9LCJ7;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Dextranucrase.  
GN Name=dexT;  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=NRRL B-512F;  
RX MEDLINE=20169623; PubMed=10705445;  
RA Funane K., Mizuno K., Takahara H., Kobayashi M.;  
RT "Gene encoding a dextranucrase-like protein in Leuconostoc  
RL Biosci. Biotechnol. Biochem. 64:29-38(2000).  
DR EMBL; AB020020; BAA90527.1; -; Genomic\_DNA.  
DR HSSP; P06278; 1VJS.  
DR GO; GO:0009250; P-glucan biosynthesis; IEA.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
SQ SEQUENCE 1016 AA; 110343 MW; 8896EFDE13CCB47 CRC64;

Query Match 100.0%; Score 108; DB 2; Length 1016;  
Best Local Similarity 100.0%; Pred. No. 3.8e-06;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVDALLQI 22

|||||  
509 DANFDSIRVDVNDVDALLQI 530

RESULT 2

Q69A94 LEUME  
ID Q69A94 LEUME PRELIMINARY; PRT; 1454 AA.  
AC Q69A94;  
DT 25-OCT-2004 (TrEMBLrel. 28, Created)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Dextranucrase (EC 2.4.1.5).  
GN Name=dexP;  
OS Leuconostoc mesenteroides  
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=IBT-PQ;  
RA Fernandez-Vazquez J.L., Lopez-Munguia A., Olvera C.;  
RT "Molecular characterization of a dextranucrase gene from Leuconostoc  
mesenteroides IBT-PQ isolated from pulque.";

```

RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY504865; AAS79426.1; -; Genomic DNA.
DR GO; GO:0047849; F:dextranase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR DR InterPro; IPR002479; CW binding.
DR DR InterPro; IPR003318; Glyco_hydro_70.
DR DR Pfam; PF01473; CW binding 1; 1.
DR DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1454 AA; 160569 MW; 2DC7342963B6FD87 CRC64;

Query Match 100.0%; Score 108; DB 2; Length 1454;
Best Local Similarity 100.0%; Pred. No. 5.4e-06;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22
Db 464 DANFDSIRVDVNDVADLLQI 485

RESULT 3
GTFC_STRMU
ID GTFC_STRMU STANDARD; PRT; 1455 AA.
AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glycosyltransferase-SI precursor (SC 2.4.1.5) (GTF-SI)
DE (Dextranucrase) (Sucrose 6-glucosyltransferase).
GN Name:gtfC; OrderedLocusNames=SMU.1005;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=89137980; PubMed=2976010; DOI=10.1016/0378-1119(88)90382-4;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";
RL Gene 69:101-109(1988).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,
MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UAI59 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental
pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE OF 1-349.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
play a key role in the development of the dental plaque because of
their ability to adhere to smooth surfaces and mediate the

```

```

CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-
fructose + (1,6-alpha-D-glucosyl)(n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-1 synthesizes water-insoluble glucans (alpha
1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 5 cell wall binding repeats.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
CC EMBL; M22054; AAA88592.1; -; Genomic DNA.
DR EMBL; D88652; BAA26102.1; -; Genomic DNA.
DR EMBL; D88655; BAA26106.1; -; Genomic DNA.
DR EMBL; D88658; BAA26110.1; -; Genomic DNA.
DR EMBL; D88661; BAA26114.1; -; Genomic DNA.
DR EMBL; D89978; BAA26120.1; -; Genomic DNA.
DR EMBL; AE014940; AAN58706.1; -; Genomic DNA.
DR PIR; JTO345; JTO345.
DR HSP; P06653; LH8G.
DR InterPro; IPR002479; Cell_wall_bd_put.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
KW Transferase.
FT SIGNAL 1 34
FT CHAIN 35 1455 Glucosyltransferase-SI.
FT REPEAT 1126 1159 A repeat.
FT REPEAT 1169 1200 A repeat.
FT REPEAT 1227 1238 C repeat.
FT REPEAT 1253 1303 AC repeat.
FT REPEAT 1318 1330 A repeat (incomplete).
FT REGION 35 1050 Catalytic (approximate).
FT REGION 1126 1455 2.4 A, 1 C and 1 AC repeats.
FT REGION 1126 1455 Glucan-binding (approximate).
FT VARIANT 21 21 V -> I (in strain GS-5).
FT VARIANT 81 81 P -> L (in strain MT4239).
FT VARIANT 106 106 D -> V (in strain GS-5).
FT VARIANT 116 116 S -> A (in strain GS-5 and strain
MT4467).
FT VARIANT 126 126 A -> T (in strain GS-5).
FT VARIANT 150 151 SR -> PK (in strain GS-5, strain MT4239
and strain MT4467).
FT VARIANT 256 256 A -> V (in strain GS-5 and strain
MT4467).
FT VARIANT 425 425 R -> N (in strain MT4251).
FT VARIANT 519 519 Y -> D (in strain MT4245 and strain
MT4251).
FT VARIANT 538 538 R -> K (in strain MT4245 and strain
MT4251).
FT VARIANT 545 545 Y -> F (in strain MT4245 and strain
MT4251).
FT VARIANT 597 597 N -> D (in strain MT4245, strain MT4251,
strain MT4467 and strain MT8148).
FT VARIANT 600 600 R -> K (in strain MT4245, strain MT4251,
strain MT4467 and strain MT8148).
FT VARIANT 601 601 A -> T (in strain GS-5).
FT VARIANT 614 614 M -> T (in strain GS-5).
FT VARIANT 727 727 T -> I (in strain MT8148).
FT VARIANT 734 734 A -> V (in strain MT8148).
FT VARIANT 964 964 L -> F (in strain MT4239).
FT VARIANT 1113 1113 N -> Y (in strain MT4239).
FT VARIANT 1118 1118 A -> T (in strain MT4239).
FT VARIANT 1204 1204 I -> V (in strain GS-5, strain MT4239,
strain MT4467 and strain MT8148).

```

```

FT VARIANT 1208 1208 V -> I (in strain MT8148).
FT VARIANT 1292 1294 DGH -> NGY (in strain GS-5, strain MT4467
FT VARIANT 1305 1369 Missing (in strain MT8148).
FT VARIANT 1326 1326 I -> V (in strain MT8148).
FT VARIANT 1331 1331 T -> A (in strain GS-5, strain MT4239,
FT VARIANT 1331 1331 strain MT4467 and strain MT8148).
FT VARIANT 1377 1377 R -> K (in strain MT8148).
FT VARIANT 1398 1398 V -> I (in strain MT8148).
FT VARIANT 1424 1424 D -> N (in strain MT4239).
FT VARIANT 1439 1439 V -> I (in strain MT4239 and strain
FT VARIANT 1444 1444 MT8148).
FT VARIANT 1444 1444 S -> P (in strain MT8148).
FT CONFLICT 1337 1455 QRLFFKSGVQAGELITERGRIKYDYPNCSNEVRNRYVR
FT TSSGNYFYGDGYALIGWVGRYVFDENGVIYASHD
FT QRHWYDYRDRGSSSARFRHSRNGFFDNFFRF ->
FT HASILSLMVFRLRESSLQSVKVSNTMILIPKMFIVM
FT (in Ref. 1).
SQ SEQUENCE 1455 AA; 162966 MW; 3CB455A99A4FEC86 CRC64;

Query Match 100.0%; Score 108; DB 1; Length 1455;
Best Local Similarity 100.0%; Pred. No. 5.4e-06;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22
DB 468 DANFDSIRVDVNDVADLLQI 489

RESULT 4
GTFB_STRMU STANDARD; PRT; 1476 AA.
AC P08987; O69384; O69387; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfB; OrderedLocusNames=SMU.1004;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.:
RL "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kencon S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of

```

```

CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes both
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 10 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC ENBL; M17361; AAA98598.1; -; Genomic_DNA.
CC ENBL; D86651; BAA26101.1; -; Genomic_DNA.
CC ENBL; D86654; BAA26105.1; -; Genomic_DNA.
CC ENBL; D86657; BAA26109.1; -; Genomic_DNA.
CC ENBL; D86660; BAA26113.1; -; Genomic_DNA.
CC ENBL; D89977; BAA26119.1; -; Genomic_DNA.
CC ENBL; AS014940; AAN58705.1; -; Genomic_DNA.
CC PIR; B33135; B33135.
CC HSSP; P06653; 1H8G.
CC InterPro; IPR002479; Cell wall bd put.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_I; 4.
CC Pfam; PF02324; Glyco_hydro_70; 1.
CC Complete proteome; Dental caries; Glycosyltransferase; Repeat; Signal;
CC Transferrase.
CC
CC SIGNAL 1 34 Potential.
CC CHAIN 35 1476 Glucosyltransferase-I.
CC REPEAT 1097 1130 A repeat..
CC REPEAT 1161 1210 1.
CC REPEAT 1225 1275 2.
CC REPEAT 1290 1340 3.
CC REPEAT 1355 1405 4.
CC REPEAT 1420 1470 5.
CC REGION 35 1051 Catalytic (approximate).
CC REGION 1097 1476 Glucan-binding (approximate).
CC REGION 1161 1470 S X tandem repeats.
CC VARIANT 62 62 S -> T (in strain MT4239).
CC VARIANT 65 65 T -> I (in strain GS-5).
CC VARIANT 68 68 V -> A (in strain GS-5, strain MT4245,
CC strain MT4251, strain MT4467 and strain
CC MT8148).
CC VARIANT 78 78 Q -> P (in strain MT4251).
CC VARIANT 86 86 I -> S (in strain GS-5, strain MT4245,
CC strain MT4251, strain MT4467 and strain
CC MT8148).
CC VARIANT 89 89 S -> F (in strain MT4251).
CC VARIANT 168 168 K -> N (in strain MT4251).
CC VARIANT 276 276 S -> D (in strain GS-5, strain MT4467 and
CC strain MT8148).
CC VARIANT 399 399 N -> R (in strain MT4239).
CC VARIANT 474 474 I -> T (in strain MT4239).
CC VARIANT 512 512 K -> R (in strain MT8148).
CC VARIANT 519 519 F -> Y (in strain MT8148).
CC VARIANT 701 701 T -> I (in strain MT8148).
CC VARIANT 708 708 A -> V (in strain MT8148).
CC VARIANT 938 938 F -> L (in strain MT8148).
CC VARIANT 952 957 FGPVPE -> YGPVPA (in strain GS-5, strain
CC MT4239 and strain MT4467).
CC VARIANT 963 964 SV -> NT (in strain GS-5, strain MT4239
CC and strain MT4467).
CC VARIANT 968 970 ADS -> VDG (in strain GS-5, strain MT4239
CC and strain MT4467).
CC VARIANT 1086 1086 A -> T (in strain MT4239).
CC VARIANT 1158 1158 S -> N (in strain MT4239).
CC VARIANT 1163 1163 H -> Y (in strain MT4251).

```

FT VARIANT 1168 1168 E -> K (in strain MT8148).  
 FT VARIANT 1182 1182 Y -> C (in strain MT8148).  
 FT VARIANT 1234 1234 A -> P (in strain MT4239).  
 FT VARIANT 1263 1263 R -> H (in strain GS-5 and strain MT4467).  
 FT VARIANT 1263 1263 Y -> P (in strain MT8148).  
 FT VARIANT 1264 1264 Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).  
 FT VARIANT 1272 1272 S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).  
 FT VARIANT 1329 1329 H -> Y (in strain GS-5 and strain MT4467).  
 FT VARIANT 1394 1394 Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).  
 FT VARIANT 1402 1402 S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).  
 FT VARIANT 1459 1459 Y -> H (in strain MT4467).  
 FT CONFLICT 570 570 R -> A (in Ref. 1).  
 FT CONFLICT 800 817 ADDVRVAASTAPSTDGK -> LKMFALRLARPHQOMA (in Ref. 1).  
 FT CONFLICT 1310 1310 H -> L (in Ref. 1).  
 SQ SEQUENCE 1476 AA; 165847 MW; 9C6E09F731B4CBCF CRC64;

Query Match 100.0%; Score 108; DB 1; Length 1476;  
 Best Local Similarity 100.0%; Pred. No. 5.5e-06;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQI 22  
 DB 442 DANFDSIRVDVNDVNDADLLQI 463

RESULT 5  
 Q55263\_9STRE PRELIMINARY; PRT; 1590 AA.  
 AC Q55263;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
 DE GTF-I.  
 GN Name=Glucosyltransferase;  
 OS Streptococcus sobrinus.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1310;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=ATCC 33478;  
 RA Sato S.;  
 RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase produced from Streptococcus sobrinus ATCC 33478.";  
 RL Kagoshima Daigaku Shigakubu Kiyo 16:23-29(1996).  
 RN [2]

PROTEIN SEQUENCE.  
 RX MEDLINE=91224988; PubMed=1827439;  
 RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;  
 RT "Isolation and sequence of an active-site peptide containing a catalytic aspartic acid from two Streptococcus sobrinus alpha-glucosyltransferases.";  
 RL J. Biol. Chem. 266:8916-8922(1991).  
 DR EMBL; D63570; BAA09792.1; -; Genomic\_DNA.  
 DR PIR; A39841; A39841.  
 DR HSSP; P06653; 1GVM.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 3.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6E4FD43 CRC64;  
 Query Match 100.0%; Score 108; DB 2; Length 1590;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-06;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQI 22  
 DB 438 DANFDSIRVDVNDVNDADLLQI 459

RESULT 6  
 Q59983\_9STRE PRELIMINARY; PRT; 1590 AA.  
 ID Q59983;  
 AC Q59983;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)  
 DE Glucosyltransferase-I precursor (EC 2.4.1.5).  
 GN Name=gtfI;  
 OS Streptococcus sobrinus.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1310;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=OMZ176;  
 RX MEDLINE=94146405; PubMed=8312602;  
 RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;  
 RT "DNA sequence of the glucosyltransferase gene of serotype d Streptococcus sobrinus.";  
 RL DNA Seq. 4:19-27(1993).  
 RN [2]  
 RP PROTEIN SEQUENCE.  
 RX MEDLINE=91224988; PubMed=1827439;  
 RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;  
 RT "Isolation and sequence of an active-site peptide containing a catalytic aspartic acid from two Streptococcus sobrinus alpha-glucosyltransferases.";  
 RL J. Biol. Chem. 266:8916-8922(1991).  
 DR EMBL; D13858; BAA02976.1; -; Genomic\_DNA.  
 DR PIR; A39841; A39841.  
 DR HSSP; P06653; 1HCX.  
 DR GO; GO:0047849; F:dextranucrase activity; IEA.  
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 2.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Glycosyltransferase; Signal; Transferase.  
 FT SIGNAL 1 38 Potential.  
 FT CHAIN 39 1590 Glucosyltransferase-I.  
 SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 100.0%; Score 108; DB 2; Length 1590;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-06;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQI 22  
 DB 438 DANFDSIRVDVNDVNDADLLQI 459

RESULT 7  
 GTFI\_STRDO STANDARD; PRT; 1597 AA.  
 ID GTFI\_STRDO  
 AC P11001;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 13-SEP-2005 (Rel. 48, Last annotation update)  
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)  
 DE (Sucrose 6-glucosyltransferase).  
 GN Name=gtfI;  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1317;

```

[1]
RN NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=MFE28;
RX MEDLINE=97308014; Pubmed=3040686;
RA Farretti J.J., Gilpin M.L., Russell R.R.B.;
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
sobrinus MFE28.";
RL J. Bacteriol. 169:4271-4278(1987).
CC -1- FUNCTION: Production of extracellular glucans, that are thought to
play a key role in the development of the dental plaque because of
their ability to adhere to smooth surfaces and mediate the
aggregation of bacterial cells and food debris.
CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-
fructose + (1,6-alpha-D-glucosyl)(n+1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
forms of glucans.
CC -1- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.
CC -1- SIMILARITY: Contains 19 cell wall binding repeats.
-----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
-----
EMBL: M17391; AAC63063.1; -; Genomic DNA.
DR InterPro; IPR002479; Cell wall bd put.
DR Pfam; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Dental caries; Glucosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38
FT CHAIN 39 1597
FT REPEAT 1099 1132 A repeat.
FT REPEAT 1163 1213 AC repeat.
FT REPEAT 1227 1277 AC repeat.
FT REPEAT 1292 1342 AC repeat.
FT REPEAT 1352 1399 B repeat.
FT REPEAT 1406 1455 AC repeat.
FT REPEAT 1465 1512 AC repeat.
FT REPEAT 1519 1568 AC repeat.
FT REGION 1582 1597 A repeat (incomplete).
FT REGION 1099 1597 Catalytic (approximate).
FT REGION 1099 1597 1-25 A, 2 B and 5 AC repeats.
FT REGION 1099 1597 Glucan-binding (approximate).
SQ SEQUENCE 1597 AA; 177080 MW; B9B86A200868798E CRC64;

Query Match 100.0%; Score 108; DB 1; Length 1597;
Best Local Similarity 100.0%; Pred. No. 5,9e-06;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQI 22
DB 444 DANFDSIRVDVNDVNDADLLQI 465

RESULT 8
Q6TXV4 LEUME
ID Q6TXV4 LEUME PRELIMINARY; PRT; 1522 AA.
AC Q6TXV4;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Dextranucrase (EC 2.4.1.5).
GN Name=dsrX;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.

[1]
RN NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=L0309;
RA Wang J.-H., Shao Y.-C., Teng D., Yang Y.-L., Zhang F.;
RT "Cloning and sequence analysis of gene coding for dextranucrase from
Leuconostoc mesenteroides L0309.";
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
EMBL: AV743959; AAQ98615.2; -; Genomic DNA.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glucosyltransferase; Transferase.
SQ SEQUENCE 1522 AA; 169103 MW; 01BCC15468B913AE CRC64;

Query Match 96.3%; Score 104; DB 2; Length 1522;
Best Local Similarity 95.5%; Pred. No. 2e-05;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQI 22
DB 537 DANFDSIRVDVNDVNDADLLQI 558

RESULT 9
Q8KRE1 LEUME
ID Q8KRE1 LEUME PRELIMINARY; PRT; 1527 AA.
AC Q8KRE1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Dextranucrase DsrD (EC 2.4.1.5).
GN Name=dsrD;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22573396; Pubmed=12686639; DOI=10.1099/mic.0.26029-0;
RA Neubauer H., Bauche A., Mollet B.;
RT "Molecular characterization and expression analysis of the
dextranucrase DsrD of Leuconostoc mesenteroides Lcc4 in homologous
RT and heterologous Lactococcus lactis cultures.";
RL Microbiology 149:973-982(2003).
DR EMBL; AY017384; AAC61158.1; -; Genomic DNA.
DR HSP; P06653; 1H8G.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glucosyltransferase; Transferase.
SQ SEQUENCE 1527 AA; 169835 MW; F9D0DE220BD89668 CRC64;

Query Match 96.3%; Score 104; DB 2; Length 1527;
Best Local Similarity 95.5%; Pred. No. 2e-05;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQI 22
DB 542 DANFDSIRVDVNDVNDADLLQI 563

RESULT 10
Q9ZAR4 LEUME
ID Q9ZAR4 LEUME PRELIMINARY; PRT; 1527 AA.
AC Q9ZAR4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

```

RC STRAIN=L0309;
RA Wang J.-H., Shao Y.-C., Teng D., Yang Y.-L., Zhang F.;
RT "Cloning and sequence analysis of gene coding for dextranucrase from
Leuconostoc mesenteroides L0309.";
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
EMBL: AV743959; AAQ98615.2; -; Genomic DNA.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glucosyltransferase; Transferase.
SQ SEQUENCE 1522 AA; 169103 MW; 01BCC15468B913AE CRC64;

Query Match 96.3%; Score 104; DB 2; Length 1522;
Best Local Similarity 95.5%; Pred. No. 2e-05;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQI 22
DB 537 DANFDSIRVDVNDVNDADLLQI 558

RESULT 9
Q8KRE1 LEUME
ID Q8KRE1 LEUME PRELIMINARY; PRT; 1527 AA.
AC Q8KRE1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Dextranucrase DsrD (EC 2.4.1.5).
GN Name=dsrD;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22573396; Pubmed=12686639; DOI=10.1099/mic.0.26029-0;
RA Neubauer H., Bauche A., Mollet B.;
RT "Molecular characterization and expression analysis of the
dextranucrase DsrD of Leuconostoc mesenteroides Lcc4 in homologous
RT and heterologous Lactococcus lactis cultures.";
RL Microbiology 149:973-982(2003).
DR EMBL; AY017384; AAC61158.1; -; Genomic DNA.
DR HSP; P06653; 1H8G.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glucosyltransferase; Transferase.
SQ SEQUENCE 1527 AA; 169835 MW; F9D0DE220BD89668 CRC64;

Query Match 96.3%; Score 104; DB 2; Length 1527;
Best Local Similarity 95.5%; Pred. No. 2e-05;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQI 22
DB 542 DANFDSIRVDVNDVNDADLLQI 563

RESULT 10
Q9ZAR4 LEUME
ID Q9ZAR4 LEUME PRELIMINARY; PRT; 1527 AA.
AC Q9ZAR4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

DE Dextranucrase.  
GN Name=DEX;  
OS Leuconostoc mesenteroides.  
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
OX NCBI\_TaxID=1245;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=NRRL B-512-F;  
RA Bhatnagar R., Singh D.K.S.;  
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U81374; RAD10952.1; -; Genomic\_DNA.  
DR HSSP; P06653; IH8G.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
DR InterPro; IPR002479; CW binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 2.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
SQ SEQUENCE 1527 AA; 169705 MW; 1DFAFA237C743398 CRC64;

Query Match 96.3%; Score 104; DB 2; Length 1527;  
Best Local Similarity 95.5%; Pred. No. 2e-05; Indels 0; Gaps 0;  
Matches 21; Conservative 0; Mismatches 1;

QY 1 DANFDSIRVDVNDVADLLQI 22  
|||||:|||||:|||||:|||||:  
DB 542 DANFDGIRVDVNDVADLLQI 563

RESULT 11  
Q5SBN0 LACRE  
ID Q5SBN0\_LACRE PRELIMINARY; PRT; 1772 AA.  
AC Q5SBN0;  
DT 01-FEB-2005 (TrEMBLrel. 29, Created)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)  
DE Glucanucrase (EC 2.4.1.5).  
OS Lactobacillus reuteri.  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=1598;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ML1;  
RX PubMed=15528655; DOI=10.1099/mic.0.27321-0;  
RA Kralj S., van Geel-Schutten G.H., Dondorff M.M.G., Kirsanovs S.,  
van der Maarel M.J.E.C., Dijkhuizen L.;  
RT "Glucan synthesis in the genus Lactobacillus: isolation and  
characterization of glucanucrase genes, enzymes and glucan products  
from six different strains.";  
RL Microbiology 150:3681-3690(2004).  
DR EMBL; AY697431; AAU08004.1; -; Genomic DNA.  
DR GO; GO:0047849; F:dextranucrase activity; IEA.  
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.  
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.  
DR GO; GO:0030245; P:cellulose catabolism; IEA.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
DR GO; GO:0000272; P:polysaccharide catabolism; IEA.  
DR InterPro; IPR002479; CW binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 2.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Glycosyltransferase; Transferase.  
SQ SEQUENCE 1772 AA; 196221 MW; 4A553F1BAD5A210B CRC64;

Query Match 95.4%; Score 103; DB 2; Length 1772;  
Best Local Similarity 95.5%; Pred. No. 3.2e-05;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22  
|||||:|||||:|||||:|||||:  
DB 1016 DANFDSIRVDVNDVADLLDI 1037

RESULT 12  
Q5SBN6 LACFE  
ID Q5SBN6\_LACFE PRELIMINARY; PRT; 1463 AA.  
AC Q5SBN6;  
DT 01-FEB-2005 (TrEMBLrel. 29, Created)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)  
DE Glucanucrase (EC 2.4.1.5).  
OS Lactobacillus fermentum.  
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
OC Lactobacillus.  
OX NCBI\_TaxID=1613;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=Kg3;  
RX PubMed=15528655; DOI=10.1099/mic.0.27321-0;  
RA Kralj S., van Geel-Schutten G.H., Dondorff M.M.G., Kirsanovs S.,  
van der Maarel M.J.E.C., Dijkhuizen L.;  
RT "Glucan synthesis in the genus Lactobacillus: isolation and  
characterization of glucanucrase genes, enzymes and glucan products  
from six different strains.";  
RL Microbiology 150:3681-3690(2004).  
DR EMBL; AY697433; AAU08008.1; -; Genomic DNA.  
DR GO; GO:0047849; F:dextranucrase activity; IEA.  
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
KW Glycosyltransferase; Transferase.  
SQ SEQUENCE 1463 AA; 161049 MW; F7EAA56AB1A3A68 CRC64;

Query Match 94.4%; Score 102; DB 2; Length 1463;  
Best Local Similarity 86.4%; Pred. No. 3.6e-05;  
Matches 19; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22  
|||||:|||||:|||||:|||||:  
DB 563 DANFDAIRVDVNDVADLLQL 584

RESULT 13  
Q00600 STRSL  
ID Q00600\_STRSL PRELIMINARY; PRT; 1518 AA.  
AC Q00600;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DE Glucosyltransferase-I.  
GN Names=gtfJ;  
OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1304;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC 25975;  
RX MEDLINE=92148377; PubMed=1838391;  
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;  
RT "Molecular characterization of a cluster of at least tow  
glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";  
J. Gen. Microbiol. 137:2577-2593(1991).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC 25975;  
RA Jacques N.;  
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
DR EMBL; Z11873; CAA77900.1; -; Genomic DNA.  
DR EMBL; M64111; AAA26896.1; -; Genomic\_DNA.  
DR PIR; A44811; A44811.  
DR HSSP; P06653; IGVM.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.



```

RT glucosyltransferase that synthesizes soluble glucans." ;
RL Infect. Immun. 58:2452-2458(1990).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-
CC fructose + (1,6-alpha-D-glucosyl)(n+1).
CC -!- ENZYME REGULATION: Glucan synthesis by GTF-S is independent of
CC primer glucan unlike GTF-I.
CC -!- MISCELLANEOUS: Synthesizes water-soluble glucans (alpha 1,6-
CC glucose).
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 10 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; M30943; AAA26898.1; -; Genomic_DNA.
CC DR HSSP; P06653; 1H8G.
CC DR InterPro; IPR002479; Cell wall bd put.
CC DR InterPro; IPR003318; Glyco_hydro_70.
CC DR Pfam; PF01473; CW_binding_1; 2.
CC DR Pfam; PF02324; Glyco_hydro_70; 1.
CC KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
CC FT SIGNAL 1 36 Or 37 (Potential).
CC FT CHAIN 37 1365 Glucosyltransferase-S.
CC FT REPEAT 157 177 Cell wall binding 1.
CC FT REPEAT 178 197 Cell wall binding 2.
CC FT REPEAT 1062 1082 Cell wall binding 3.
CC FT REPEAT 1083 1102 Cell wall binding 4.
CC FT REPEAT 1150 1169 Cell wall binding 5.
CC FT REPEAT 1170 1190 Cell wall binding 6.
CC FT REPEAT 1225 1243 Cell wall binding 7.
CC FT REPEAT 1289 1308 Cell wall binding 8.
CC FT REPEAT 1309 1328 Cell wall binding 9.
CC FT REPEAT 1331 1352 Cell wall binding 10.
CC FT REGION 198 1061 Catalytic (approximate).
CC SQ SEQUENCE 1365 AA; 151591 MW, 167296B5A2E8C476 CRC64;

Query Match 90.7%; Score 98; DB 1; Length 1365;
Best Local Similarity 86.4%; Pred. No. 0.00012;
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22
| | | | : | | | | | | | | | |
Db 428 DANFDGVRVDVNDVNADLLQI 449

Search completed: February 10, 2006, 22:31:11
Job time : 15.1028 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:31:39 ; Search time 3.96976 Seconds  
(without alignments)  
458.180 Million cell updates/sec

Title: US-10-797-821-25

Perfect score: 108

Sequence: 1 DANFDSIRVDVNDVADLLQI 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	100.0	22	1	US-08-057-162B-1
2	108	100.0	1375	2	US-09-210-361-4
3	108	100.0	1375	2	US-09-740-274-4
4	108	100.0	1475	2	US-09-007-999-2
5	108	100.0	1475	2	US-09-210-361-2
6	108	100.0	1475	2	US-09-740-274-2
7	104	96.3	522	2	US-09-995-749A-11
8	104	96.3	523	2	US-09-604-957-5
9	96	88.9	1577	1	US-08-793-824-2
10	95	88.0	584	2	US-09-604-957-6
11	95	88.0	584	2	US-09-995-749A-12
12	95	88.0	2057	2	US-09-499-203-2
13	94	87.0	545	2	US-09-604-957-4
14	94	87.0	545	2	US-09-995-749A-10
15	94	87.0	1430	2	US-09-008-172-2
16	94	87.0	1430	2	US-09-210-361-6
17	94	87.0	1430	2	US-09-740-274-6
18	87	80.6	535	2	US-09-604-957-7
19	87	80.6	535	2	US-09-995-749A-13
20	87	80.6	1278	2	US-09-604-957-3
21	87	80.6	1781	2	US-09-995-749A-2
22	54	50.0	624	2	US-08-947-965-78
23	54	50.0	655	1	US-08-469-202-27
24	54	50.0	655	1	US-08-469-202-28
25	54	50.0	655	1	US-08-484-434C-34
26	54	50.0	655	1	US-08-484-434C-35
27	54	50.0	655	2	US-09-384-361-34

28	54	50.0	655	2	US-09-384-361-35	Sequence 35, Appli
29	46	42.6	25	2	US-08-888-080A-6	Sequence 6, Appli
30	46	42.6	25	2	US-09-577-027-6	Sequence 6, Appli
31	46	42.6	38	2	US-08-888-080A-7	Sequence 7, Appli
32	46	42.6	38	2	US-08-888-080A-8	Sequence 8, Appli
33	46	42.6	38	2	US-09-577-027-7	Sequence 7, Appli
34	46	42.6	38	2	US-09-577-027-8	Sequence 8, Appli
35	46	42.6	740	2	US-09-248-796A-17732	Sequence 17732, A
36	46	42.6	762	2	US-09-786-480B-25	Sequence 25, Appli
37	45	41.7	465	2	US-09-252-991A-21721	Sequence 21721, A
38	45	41.7	793	2	US-09-489-039A-7449	Sequence 7449, Ap
39	44	40.7	25	2	US-08-888-080A-4	Sequence 4, Appli
40	44	40.7	25	2	US-09-577-027-4	Sequence 4, Appli
41	44	40.7	209	2	US-09-583-110-3039	Sequence 3039, Ap
42	44	40.7	217	2	US-09-107-433-4632	Sequence 4632, Ap
43	43.5	40.3	3177	1	US-08-477-451-4	Sequence 4, Appli
44	43	39.8	330	2	US-09-252-991A-24853	Sequence 24853, A
45	43	39.8	668	2	US-09-538-092-599	Sequence 599, App

ALIGNMENTS

RESULT 1  
US-08-057-162B-1  
; Sequence 1, Application US/08057162B  
; Patent No. 5686075  
; GENERAL INFORMATION:  
; APPLICANT: Taubman, Martin A.  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR DENTAL CARIES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/057,162B  
; FILING DATE: 30-APR-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/877,295  
; FILING DATE: 01-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wagner, Richard W.  
; REGISTRATION NUMBER: 34,480  
; REFERENCE/DOCKET NUMBER: FDC92-01A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 22 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-057-162B-1

Query Match 100.0%; Score 108; DB 1; Length 22;  
Best Local Similarity 100.0%; Pred. No. 3,9e-11;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22

|||||

```
Db 1 DANFDSIRVDVNDVADLLQI 22
US-09-740-274-4
Query Match 100.0%; Score 108; DB 2; Length 1375;
Best Local Similarity 100.0%; Pred. No. 6.4e-09;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match 100.0%; Score 108; DB 2; Length 1375;
Best Local Similarity 100.0%; Pred. No. 6.4e-09;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22
Db 468 DANFDSIRVDVNDVADLLQI 489

RESULT 3
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans

US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match 100.0%; Score 108; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 7e-09;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22
Db 442 DANFDSIRVDVNDVADLLQI 463

RESULT 5
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2

Query Match 100.0%; Score 108; DB 2; Length 1475;
```

```
Best Local Similarity 100.0%; Pred. No. 7e-09; Score 104; DB 2; Length 1475;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVAVNDVADLLQI 22
Db 442 DANFDSIRVDVAVNDVADLLQI 463

RESULT 6
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Fast-SEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 100.0%; Score 108; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 7e-09;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVAVNDVADLLQI 22
Db 442 DANFDSIRVDVAVNDVADLLQI 463

RESULT 7
US-09-995-749A-11
; Sequence 11, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAQUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749A-11

Best Local Similarity 96.3%; Score 104; DB 2; Length 523;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVAVNDVADLLQI 22
Db 36 DANFDGIRVDVAVNDVADLLQI 57

RESULT 8
US-09-604-957-5
; Sequence 5, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAQUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-604-957-5

Query Match 96.3%; Score 104; DB 2; Length 523;
Best Local Similarity 95.5%; Pred. No. 8.8e-09;
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVAVNDVADLLQI 22
Db 36 DANFDGIRVDVAVNDVADLLQI 57

RESULT 9
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to
; TITLE OF INVENTION: Increase Stored Carbohydrates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Griffith Hack & Co
; STREET: Level 8, 168 Walker Street
; CITY: No. 5981838th Sydney
; STATE: New South Wales
; COUNTRY: Australia
; ZIP: 2060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/793,824
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PM7643
; FILING DATE: 24-AUG-1994
; TELECOMMUNICATION INFORMATION:
```

```
; TELEPHONE: 61 2 9957 5944
; TELEFAX: 61 2 957 6288
; TELEX: 26547
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1577 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus salivarius
; US-08-793-824-2

Query Match      88.9%; Score 96; DB 1; Length 1577;
Best Local Similarity 85.7%; Pred. No. 7e-07;
Matches 18; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 ANFDSIRVDAVNDVNDADLLQI 22
Db 550 ANFDGVRIDA VNDVNDADLLQI 570

RESULT 10
US-09-604-957-6
; Sequence 6, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
; US-09-604-957-6

Query Match      88.0%; Score 95; DB 2; Length 584;
Best Local Similarity 86.4%; Pred. No. 3e-07;
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDAVNDVNDADLLQI 22
Db 626 NANFDGIRVDA VNDVNDADLLKI 647

RESULT 11
US-09-995-749A-12
; Sequence 12, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
; US-09-995-749A-12

Query Match      88.0%; Score 95; DB 2; Length 584;
Best Local Similarity 86.4%; Pred. No. 3e-07;
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDAVNDVNDADLLQI 22
Db 626 NANFDGIRVDA VNDVNDADLLKI 647

RESULT 12
US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-198P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
; US-09-499-203-2

Query Match      88.0%; Score 95; DB 2; Length 2057;
Best Local Similarity 86.4%; Pred. No. 1.4e-06;
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDAVNDVNDADLLQI 22
Db 626 NANFDGIRVDA VNDVNDADLLKI 647

RESULT 13
US-09-604-957-4
; Sequence 4, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
; US-09-604-957-4

Query Match      87.0%; Score 94; DB 2; Length 545;
Best Local Similarity 81.8%; Pred. No. 4e-07;
Matches 18; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

QY 1 DANFDSIRVDVNDVADLLQI 22  
Db 36 EAFDGVVRVDVNDVADLLQI 57

## RESULT 14

US-09-995-749A-10  
; Sequence 10, Application US/09995749A  
; Patent No. 6867026  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
; FILE REFERENCE: B043388-CIP  
; CURRENT APPLICATION NUMBER: US/09/995,749A  
; CURRENT FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: 09/604,957  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: EPO 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 545  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-995-749A-10

Query Match 87.0%; Score 94; DB 2; Length 545;  
Best Local Similarity 81.8%; Pred. No. 4e-07;  
Matches 18; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22  
Db 36 EAFDGVVRVDVNDVADLLQI 57

## RESULT 15

US-09-008-172-2  
; Sequence 2, Application US/09008172  
; Patent No. 6127602  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 03589  
; CURRENT APPLICATION NUMBER: US/09/008,172  
; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-008-172-2

Query Match 87.0%; Score 94; DB 2; Length 1430;  
Best Local Similarity 81.8%; Pred. No. 1.3e-06;  
Matches 18; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22  
Db 456 EAFDGVVRVDVNDVADLLQI 477

Search completed: February 10, 2006, 22:36:14  
Job time : 4.96976 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:14:04 ; Search time 12.9073 Seconds  
(without alignments)  
712.176 Million cell updates/sec

Title: US-10-797-821-25  
Perfect score: 108  
Sequence: 1 DANFDSIRVDVNDVADLLQI 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA>Main:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	100.0	22	4	US-10-383-930-25
2	108	100.0	22	5	US-10-797-821-25
3	108	100.0	1375	3	US-09-740-274-4
4	108	100.0	1375	4	US-10-383-930-35
5	108	100.0	1375	5	US-10-797-821-35
6	108	100.0	1475	3	US-09-740-274-2
7	108	100.0	1475	4	US-10-383-930-34
8	108	100.0	1475	5	US-10-797-821-34
9	108	100.0	1590	4	US-10-383-930-37
10	108	100.0	1590	5	US-10-797-821-37
11	104	96.3	21	4	US-10-383-930-24
12	104	96.3	21	5	US-10-797-821-24
13	104	96.3	522	3	US-09-995-749A-11
14	103	95.4	221	5	US-10-484-218-8
15	103	95.4	1771	5	US-10-484-218-14
16	102	94.4	23	2	US-08-967-573A-11
17	102	94.4	1006	5	US-10-484-218-22
18	100	92.6	1518	4	US-10-383-930-40
19	100	92.6	1518	5	US-10-797-821-40
20	98	90.7	223	5	US-10-484-218-6
21	98	90.7	1365	4	US-10-383-930-39
22	98	90.7	1365	5	US-10-797-821-39
23	97	89.8	1554	4	US-10-383-930-38
24	97	89.8	1554	5	US-10-797-821-38
25	96	88.9	1595	5	US-10-484-218-20
26	95	88.0	584	3	US-09-995-749A-12
27	95	88.0	2057	4	US-10-417-280A-2

28	94	87.0	545	3	US-09-995-749A-10	Sequence 10, Appl
29	94	87.0	1430	3	US-09-740-274-6	Sequence 6, Appl
30	94	87.0	1430	4	US-10-383-930-36	Sequence 36, Appl
31	94	87.0	1430	5	US-10-797-821-36	Sequence 36, Appl
32	93	86.1	223	5	US-10-484-218-10	Sequence 10, Appl
33	91	84.3	221	5	US-10-484-218-2	Sequence 2, Appl
34	91	84.3	525	5	US-10-484-218-23	Sequence 23, Appl
35	91	84.3	1777	5	US-10-484-218-12	Sequence 12, Appl
36	88	81.5	224	5	US-10-484-218-4	Sequence 4, Appl
37	88	81.5	1497	5	US-10-484-218-18	Sequence 18, Appl
38	87	80.6	535	3	US-09-995-749A-13	Sequence 13, Appl
39	87	80.6	1781	3	US-09-995-749A-2	Sequence 2, Appl
40	79	73.1	1787	5	US-10-484-218-16	Sequence 16, Appl
41	50	46.3	590	4	US-10-425-115-310890	Sequence 310890, A
42	46.5	43.1	762	4	US-10-369-493-17708	Sequence 17708, A
43	46	42.6	25	3	US-09-922-568-6	Sequence 6, Appl
44	46	42.6	38	3	US-09-922-568-7	Sequence 7, Appl
45	46	42.6	38	3	US-09-922-568-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1  
US-10-383-930-25  
; Sequence 25, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 25  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-25

Query Match	100.0%;	Score 108;	DB 4;	Length 22;
Best Local Similarity	100.0%;	Pred. No. 2.3e-09;		
Matches	22;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	DANFDSIRVDVNDVADLLQI	22	
Db	1	DANFDSIRVDVNDVADLLQI	22	
RESULT 2				
US-10-797-821-25				
; Sequence 25, Application US/10797821				
; Publication No. US20050031633A1				
; GENERAL INFORMATION:				
; APPLICANT: Smith, Daniel J.				
; APPLICANT: Taubman, Martin A.				
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens				
; FILE REFERENCE: 25669-020				
; CURRENT APPLICATION NUMBER: US/10/797,821				
; CURRENT FILING DATE: 2004-03-09				
; PRIOR APPLICATION NUMBER: 10/383,930				
; PRIOR FILING DATE: 2003-03-07				
; PRIOR APPLICATION NUMBER: 60/363,209				
; PRIOR FILING DATE: 2002-03-07				
; PRIOR APPLICATION NUMBER: 60/402,483				
; PRIOR FILING DATE: 2002-08-08				
; PRIOR APPLICATION NUMBER: 09/290,049				

; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 25  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: GTF-derived catalytic (CAT) peptide  
US-10-797-821-25

Query Match 100.0%; Score 108; DB 5; Length 22;  
Best Local Similarity 100.0%; Pred. No. 2.3e-09;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVDADLLQI 22  
|||||  
Db 1 DANFDSIRVDVNDVDADLLQI 22

RESULT 3  
US-09-740-274-4  
; Sequence 4, Application US/09740274  
; Patent No. US20020031826A1  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 100.0%; Score 108; DB 3; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVDADLLQI 22  
|||||  
Db 468 DANFDSIRVDVNDVDADLLQI 489

RESULT 4  
US-10-383-930-35  
; Sequence 35, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein

; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 35  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-35

Query Match 100.0%; Score 108; DB 4; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVDADLLQI 22  
|||||  
Db 468 DANFDSIRVDVNDVDADLLQI 489

RESULT 5  
US-10-797-821-35  
; Sequence 35, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 35  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-35

Query Match 100.0%; Score 108; DB 5; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 2.7e-07;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVDADLLQI 22  
|||||  
Db 468 DANFDSIRVDVNDVDADLLQI 489

RESULT 6  
US-09-740-274-2  
; Sequence 2, Application US/09740274  
; Patent No. US20020031826A1  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274

```
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match          100.0%; Score 108; DB 3; Length 1475;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DANFDSIRVDVNDVNDADLLQI 22
      |||||
Db      442 DANFDSIRVDVNDVNDADLLQI 463

RESULT 7
US-10-383-930-34
; Sequence 34, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-34

Query Match          100.0%; Score 108; DB 4; Length 1475;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DANFDSIRVDVNDVNDADLLQI 22
      |||||
Db      442 DANFDSIRVDVNDVNDADLLQI 463

RESULT 8
US-10-797-821-34
; Sequence 34, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-34

Query Match          100.0%; Score 108; DB 5; Length 1475;
Best Local Similarity 100.0%; Pred. No. 2.9e-07;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DANFDSIRVDVNDVNDADLLQI 22
      |||||
Db      442 DANFDSIRVDVNDVNDADLLQI 463

RESULT 9
US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match          100.0%; Score 108; DB 4; Length 1590;
Best Local Similarity 100.0%; Pred. No. 3.2e-07;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DANFDSIRVDVNDVNDADLLQI 22
      |||||
Db      438 DANFDSIRVDVNDVNDADLLQI 459

RESULT 10
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
```

4

; CURRENT FILING DATE: 2004-01-20  
; PRIOR APPLICATION NUMBER: PCT/NL02/00495  
; PRIOR FILING DATE: 2002-07-22  
; PRIOR APPLICATION NUMBER: EP 01202752.0  
; PRIOR FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: EP 01202841.1  
; PRIOR FILING DATE: 2001-07-25  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 8  
; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Lactobacillus sp.  
US-10-484-218-8

Query Match 95.4%; Score 103; DB 5; Length 221;  
Best Local Similarity 95.5%; Pred. No. 1.8e-07;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22  
| | | | | | | | | | | | | | | | | | | | | |  
Db 30 DANFDSIRVDVNDVADLLDI 51

RESULT 15  
US-10-484-218-14  
; Sequence 14, Application US/10484218  
; Publication No. US20050059633A1  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITINA HENDRIKA  
; TITLE OF INVENTION: GLUCANS AND GLUCANSUCRASES DERIVED FROM  
; FILE REFERENCE: 2001-1316  
; CURRENT APPLICATION NUMBER: US/10/484,218  
; CURRENT FILING DATE: 2004-01-20  
; PRIOR APPLICATION NUMBER: PCT/NL02/00495  
; PRIOR FILING DATE: 2002-07-22  
; PRIOR APPLICATION NUMBER: EP 01202752.0  
; PRIOR FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: EP 01202841.1  
; PRIOR FILING DATE: 2001-07-25  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 14  
; LENGTH: 1771  
; TYPE: PRT  
; ORGANISM: Lactobacillus reuteri  
US-10-484-218-14

Query Match 95.4%; Score 103; DB 5; Length 1771;  
Best Local Similarity 95.5%; Pred. No. 2e-06;  
Matches 21; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQI 22  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1015 DANFDSIRVDVNDVADLLDI 1036

Search completed: February 10, 2006, 23:25:10  
Job time : 12.9073 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:15:51 ; Search time 0.953629 Seconds  
(without alignments)  
302.734 Million cell updates/sec

Title: US-10-797-821-25  
Perfect score: 108  
Sequence: 1 DANFDSIRVDVNDVADLLQI 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pdb:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pdb:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pdb:\*  
4: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pdb:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pdb:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pdb:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pdb:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pdb:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42.5	39.4	772	6	US-10-858-730-77
2	42	38.9	295	6	US-10-467-657-5222
3	40.5	37.5	449	6	US-10-763-712A-73
4	40	37.0	430	6	US-10-467-657-2346
5	40	37.0	480	7	US-11-136-244-9
6	40	37.0	619	6	US-10-999-886-3
7	40	37.0	619	7	US-11-136-244-3
8	40	37.0	640	6	US-10-999-886-4
9	40	37.0	640	7	US-11-136-244-4
10	39	36.1	199	6	US-10-793-626-2546
11	39	36.1	447	6	US-10-967-527A-14
12	39	36.1	448	6	US-10-967-527A-16
13	39	36.1	610	7	US-11-055-822-1146
14	39	36.1	610	6	US-11-124-291-6
15	39	36.1	643	6	US-10-873-427A-4
16	39	36.1	737	7	US-11-128-420-9
17	39	36.1	884	6	US-10-878-556A-9
18	39	36.1	974	6	US-10-821-234-1152
19	38.5	35.6	503	7	US-11-078-991-1
20	38	35.2	237	6	US-10-510-386-34
21	38	35.2	325	7	US-11-074-176-370
22	38	35.2	332	6	US-10-793-626-2456
23	38	35.2	376	6	US-10-467-657-7330
24	38	35.2	505	6	US-10-793-626-2220
25	38	35.2	1531	7	US-11-087-227-24

ALIGNMENTS

RESULT 1  
US-10-858-730-77  
; Sequence 77, Application US/10858730  
; Publication No. US20050255568A1  
; GENERAL INFORMATION:  
; APPLICANT: Bailey, Richard B.  
; APPLICANT: Blomquist, Paul  
; APPLICANT: Doten, Reed  
; APPLICANT: Driggers, Edward M.  
; APPLICANT: Madden, Kevin T.  
; APPLICANT: O'Leary, Jessica  
; APPLICANT: O'Toole, George  
; APPLICANT: Trueheart, Joshua  
; APPLICANT: Walbridge, Michael J.  
; APPLICANT: Yorgey, Peter S.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR AMINO ACID  
; TITLE OF INVENTION: PRODUCTION  
; FILE REFERENCE: 14184-030001  
; CURRENT APPLICATION NUMBER: US/10/858,730  
; PRIOR FILING DATE: 2004-06-01  
; PRIOR APPLICATION NUMBER: US 60/475,000  
; PRIOR FILING DATE: 2003-05-30  
; PRIOR APPLICATION NUMBER: US 60/551,860  
; PRIOR FILING DATE: 2004-03-10  
; NUMBER OF SEQ ID NOS: 364  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 77  
; LENGTH: 772  
; TYPE: PRT  
; ORGANISM: Streptomyces coelicolor  
US-10-858-730-77

Query Match 39.4%; Score 42.5; DB 6; Length 772;  
Best Local Similarity 42.9%; Pred. No. 49;  
Matches 9; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Qy 2 ANFDSIRVDVNDVADLLQI 22  
Db 660 AEFEDI-VQAIDDLADVISL 679

RESULT 2  
US-10-467-657-5222  
; Sequence 5222, Application US/10467657  
; Publication No. US20050260581A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; APPLICANT: FONTANA Maria Rita

Sequence 211, App  
Sequence 1676, App  
Sequence 226, App  
Sequence 224, App  
Sequence 2, Appli  
Sequence 4, Appli  
Sequence 1, Appli  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 1358, Ap  
Sequence 8534, Ap  
Sequence 283, App  
Sequence 174, App  
Sequence 281, App  
Sequence 1116, Ap  
Sequence 2832, Ap  
Sequence 1473, Ap  
Sequence 63, Appl  
Sequence 105, App

```
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 5222
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-5222
```

```
Query Match      38.9%; Score 42; DB 6; Length 295;
Best Local Similarity 18.2%; Pred. No. 18;
Matches 4; Conservative 15; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 DANFDSIRVDVNDVNDADLLQI 22
Db      179 NSNPDAVRLDKIEHINHEIEM 200
```

```
RESULT 3
US-10-763-712A-73
; Sequence 73, Application US/10763712A
; Publication No. US20050266541A1
; GENERAL INFORMATION:
; APPLICANT: Solazyme, Inc.
; APPLICANT: Dillon, Harrison F.
; TITLE OF INVENTION: Methods and Compositions for Evolving Microbial Hydrogen
; TITLE OF INVENTION: Production
; FILE REFERENCE: H2042101-CIP
; CURRENT APPLICATION NUMBER: US/10/763,712A
; CURRENT FILING DATE: 2004-01-31
; PRIOR APPLICATION NUMBER: US 10/287,750
; PRIOR FILING DATE: 2002-11-04
; PRIOR APPLICATION NUMBER: US 10/411,910
; PRIOR FILING DATE: 2003-04-12
; PRIOR APPLICATION NUMBER: US 60/500,032
; PRIOR FILING DATE: 2003-09-03
; NUMBER OF SEQ ID NOS: 184
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 73
; LENGTH: 449
; TYPE: PRT
; ORGANISM: Clostridium perfringens
US-10-763-712A-73
```

```
Query Match      37.5%; Score 40.5; DB 6; Length 449;
Best Local Similarity 42.9%; Pred. No. 52;
Matches 9; Conservative 4; Mismatches 5; Indels 3; Gaps 1;
```

```
QY      4 FDSIRVDVNDVNDVNDADLLQI 21
Db      101 FDAITDIPIDNSKYIDPNLCQ 121
```

```
RESULT 4
US-10-467-657-2346
; Sequence 2346, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
```

```
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 2346
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-2346
```

```
Query Match      37.0%; Score 40; DB 6; Length 430;
Best Local Similarity 63.6%; Pred. No. 58;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      11 AVDNDVADLLQI 21
Db      83 AVDNDIDLIYQ 93
```

```
RESULT 5
US-11-136-244-9
; Sequence 9, Application US/11136244
; Publication No. US20060003408A1
; GENERAL INFORMATION:
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Fiske, Susan M.
; APPLICANT: Lantz, Suzanne E.
; APPLICANT: Neefe-Kruithof, Paulien
; APPLICANT: Pepsin, Michael Jay
; APPLICANT: Jayarama, Shetty K.
; TITLE OF INVENTION: Acid-Stable Alpha Amylases Having Granular Starch Hydrolyzing
; TITLE OF INVENTION: Activity and Enzyme Compositions
; FILE REFERENCE: GC873
; CURRENT APPLICATION NUMBER: US/11/136,244
; CURRENT FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US 60/575,175
; PRIOR FILING DATE: 2004-05-27
; PRIOR APPLICATION NUMBER: US 60/605,437
; PRIOR FILING DATE: 2004-08-30
; PRIOR APPLICATION NUMBER: US 60/647,925
; PRIOR FILING DATE: 2005-01-28
; PRIOR APPLICATION NUMBER: PCT/US04/40040
; PRIOR FILING DATE: 2004-11-30
; PRIOR APPLICATION NUMBER: PCT/US04/41276
; PRIOR FILING DATE: 2004-12-09
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Aspergillus kawachi
US-11-136-244-9
```

```
Query Match      37.0%; Score 40; DB 7; Length 480;
Best Local Similarity 42.9%; Pred. No. 67;
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      5 DSIRVDVNDVNDAD 18
Db      201 DGLRIDSVEVEPDP 214
```

```
RESULT 6
US-10-999-886-3
; Sequence 3, Application US/10999886
; Publication No. US20050266543A1
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; APPLICANT: Dunn-Coleman, Nigel
; APPLICANT: Lantero, Oreste
; APPLICANT: Shetty, Jayarama K.
```



; APPLICANT: Suzanne, Lantz E.  
; APPLICANT: Michael, Pepsin J.  
; TITLE OF INVENTION: Heterologous Expression of an Aspergillus kawachi Acid-Stable  
; TITLE OF INVENTION: Alpha Amylase and Applications in Granular Starch Hydrolysis  
; FILE REFERENCE: GC841  
; CURRENT APPLICATION NUMBER: US/10/999,886  
; CURRENT FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: US 60/605,437  
; PRIOR FILING DATE: 2004-08-30  
; PRIOR APPLICATION NUMBER: US 60/575,175  
; PRIOR FILING DATE: 2004-05-27  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3  
; LENGTH: 619  
; TYPE: PRT  
; ORGANISM: Aspergillus kawachi  
US-10-999-886-3

Query Match 37.0%; Score 40; DB 6; Length 619;  
Best Local Similarity 42.9%; Pred. No. 91;  
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 5 DSIRVDVNDVAD 18  
|:|:|:|:|:  
Db 201 DGLRIDSVEEVPD 214

RESULT 7  
US-11-136-244-3  
; Sequence 3, Application US/11136244  
; Publication No. US20060003408A1  
; GENERAL INFORMATION:  
; APPLICANT: Dunn-Coleman, Nigel  
; APPLICANT: Fiske, Susan M.  
; APPLICANT: Lantz, Suzanne E.  
; APPLICANT: Neefe-Kruitthof, Paulien  
; APPLICANT: Pepsin, Michael Jay  
; APPLICANT: Jayarama, Shetty K.  
; TITLE OF INVENTION: Acid-Stable Alpha Amylases Having Granular Starch Hydrolyzing  
; TITLE OF INVENTION: Activity and Enzyme Compositions  
; FILE REFERENCE: GC873  
; CURRENT APPLICATION NUMBER: US/11/136,244  
; CURRENT FILING DATE: 2005-05-24  
; PRIOR APPLICATION NUMBER: US 60/575,175  
; PRIOR FILING DATE: 2004-05-27  
; PRIOR APPLICATION NUMBER: US 60/605,437  
; PRIOR FILING DATE: 2004-08-30  
; PRIOR APPLICATION NUMBER: US 60/647,925  
; PRIOR FILING DATE: 2005-01-28  
; PRIOR APPLICATION NUMBER: PCT/US04/40040  
; PRIOR FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: PCT/US04/41276  
; PRIOR FILING DATE: 2004-12-09  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3  
; LENGTH: 619  
; TYPE: PRT  
; ORGANISM: Aspergillus kawachi  
US-11-136-244-3

Query Match 37.0%; Score 40; DB 7; Length 619;  
Best Local Similarity 42.9%; Pred. No. 91;  
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 5 DSIRVDVNDVAD 18  
|:|:|:|:|:  
Db 201 DGLRIDSVEEVPD 214

RESULT 8  
US-10-999-886-4

; Sequence 4, Application US/10999886  
; Publication No. US20050286543A1  
; GENERAL INFORMATION:  
; APPLICANT: Genencor International, Inc.  
; APPLICANT: Dunn-Coleman, Nigel  
; APPLICANT: Lantero, Oreste  
; APPLICANT: Shetty, Jayarama K.  
; APPLICANT: Suzanne, Lantz E.  
; APPLICANT: Michael, Pepsin J.  
; TITLE OF INVENTION: Heterologous Expression of an Aspergillus kawachi Acid-Stable  
; TITLE OF INVENTION: Alpha Amylase and Applications in Granular Starch Hydrolysis  
; FILE REFERENCE: GC841  
; CURRENT APPLICATION NUMBER: US/10/999,886  
; CURRENT FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: US 60/605,437  
; PRIOR FILING DATE: 2004-08-30  
; PRIOR APPLICATION NUMBER: US 60/575,175  
; PRIOR FILING DATE: 2004-05-27  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 4  
; LENGTH: 640  
; TYPE: PRT  
; ORGANISM: Aspergillus kawachi  
US-10-999-886-4

Query Match 37.0%; Score 40; DB 6; Length 640;  
Best Local Similarity 42.9%; Pred. No. 94;  
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 5 DSIRVDVNDVAD 18  
|:|:|:|:|:  
Db 222 DGLRIDSVEEVPD 235

RESULT 9  
US-11-136-244-4  
; Sequence 4, Application US/11136244  
; Publication No. US20060003408A1  
; GENERAL INFORMATION:  
; APPLICANT: Dunn-Coleman, Nigel  
; APPLICANT: Fiske, Susan M.  
; APPLICANT: Lantz, Suzanne E.  
; APPLICANT: Neefe-Kruitthof, Paulien  
; APPLICANT: Pepsin, Michael Jay  
; APPLICANT: Jayarama, Shetty K.  
; TITLE OF INVENTION: Acid-Stable Alpha Amylases Having Granular Starch Hydrolyzing  
; TITLE OF INVENTION: Activity and Enzyme Compositions  
; FILE REFERENCE: GC873  
; CURRENT APPLICATION NUMBER: US/11/136,244  
; CURRENT FILING DATE: 2005-05-24  
; PRIOR APPLICATION NUMBER: US 60/575,175  
; PRIOR FILING DATE: 2004-05-27  
; PRIOR APPLICATION NUMBER: US 60/605,437  
; PRIOR FILING DATE: 2004-08-30  
; PRIOR APPLICATION NUMBER: US 60/647,925  
; PRIOR FILING DATE: 2005-01-28  
; PRIOR APPLICATION NUMBER: PCT/US04/40040  
; PRIOR FILING DATE: 2004-11-30  
; PRIOR APPLICATION NUMBER: PCT/US04/41276  
; PRIOR FILING DATE: 2004-12-09  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 4  
; LENGTH: 640  
; TYPE: PRT  
; ORGANISM: Aspergillus kawachi  
US-11-136-244-4

Query Match 37.0%; Score 40; DB 7; Length 640;  
Best Local Similarity 42.9%; Pred. No. 94;  
Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

```
QY      5  DSRVDVNDVNDAD 18
      | :|:|:|:|:|:|
Db      222 DGLRDSVEEVEPD 235

RESULT 10
US-10-793-626-2546
; Sequence 2546, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2546
; LENGTH: 199
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2546

Query Match      36.1%; Score 39; DB 6; Length 199;
Best Local Similarity 53.8%; Pred. No. 33;
Matches 7; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY      9  VDADVNDVNDLLQ 21
      | :|:|:|:|:|:|
Db      167 IDAVENDASILR 179

RESULT 11
US-10-967-527A-14
; Sequence 14, Application US/10967527A
; Publication No. US20050256041A1
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Holloway, James L.
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: Ztnfr14, A Tumor Necrosis Factor
; FILE REFERENCE: 03-17
; CURRENT APPLICATION NUMBER: US/10/967,527A
; CURRENT FILING DATE: 2004-10-18
; PRIOR APPLICATION NUMBER: 60/511,698
; PRIOR FILING DATE: 2003-10-16
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 447
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-967-527A-14

Query Match      36.1%; Score 39; DB 6; Length 447;
Best Local Similarity 50.0%; Pred. No. 87;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY      8  RVDADVNDVNDLLQ 21
      | :|:|:|:|:|:|
Db      419 RLDAVESLCADILE 432

RESULT 12
US-10-967-527A-16
; Sequence 16, Application US/10967527A
; Publication No. US20050256041A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Fox, Brian A.
; APPLICANT: Holloway, James L.
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: Ztnfr14, A Tumor Necrosis Factor
; FILE REFERENCE: 03-17
; CURRENT APPLICATION NUMBER: US/10/967,527A
; CURRENT FILING DATE: 2004-10-18
; PRIOR APPLICATION NUMBER: 60/511,698
; PRIOR FILING DATE: 2003-10-16
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 448
; TYPE: PRT
; ORGANISM: mus musculus
US-10-967-527A-16

Query Match      36.1%; Score 39; DB 6; Length 448;
Best Local Similarity 50.0%; Pred. No. 87;
Matches 7; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

QY      8  RVDADVNDVNDLLQ 21
      | :|:|:|:|:|:|
Db      420 RLDAVESLCADILE 433

RESULT 13
US-11-055-822-1146
; Sequence 1146, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; FILE REFERENCE: BGI-121CPCN
; CURRENT APPLICATION NUMBER: US/11/055,822
; CURRENT FILING DATE: 2005-02-11
; PRIOR APPLICATION NUMBER: 09/606,740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141,031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142,101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148,613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187,970
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19930476.9
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19931415.2
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931418.7
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931419.5
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; Remaining Prior Application data removed - See File wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1158
; SEQ ID NO 1146
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-11-055-822-1146

Query Match      36.1%; Score 39; DB 7; Length 610;
Best Local Similarity 43.8%; Pred. No. 1.3e+02;
```

```
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
Qy 1 DANFDSIRVDVND 16
| : | : | : | : |
Db 276 DFHVDGLRLDAVHSLD 291

RESULT 14
US-11-124-291-6
; Sequence 6, Application US/11124291
; Publication No. US20050266536A1
; GENERAL INFORMATION:
; APPLICANT: WOLF, ANDREAS
; APPLICANT: SCHISCHKA, NATALIE
; APPLICANT: HERMANN, THOMAS
; APPLICANT: MORBACH, SUSANNE
; APPLICANT: KRAEMER, REINHARD
; TITLE OF INVENTION: PROCESS FOR THE FERMENTATIVE PREPARATION OF L-AMINO ACIDS USING
; FILE REFERENCE: 223168USOX
; CURRENT APPLICATION NUMBER: US/11/124,291
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US/10/212,219
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US 60/316,276
; PRIOR FILING DATE: 2001-09-04
; PRIOR APPLICATION NUMBER: DE 101 39 062.9
; PRIOR FILING DATE: 2001-09-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 610
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-11-124-291-6

Query Match 36.1%; Score 39; DB 7; Length 610;
Best Local Similarity 43.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
Qy 1 DANFDSIRVDVND 16
| : | : | : | : |
Db 276 DFHVDGLRLDAVHSLD 291

RESULT 15
US-10-873-427A-4
; Sequence 4, Application US/10873427A
; Publication No. US20050281914A1
; GENERAL INFORMATION:
; APPLICANT: STEELE, JAMES L.
; APPLICANT: BROADBENT, JEFFREY R.
; APPLICANT: SRIDHAR, VIDYA R.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS INVOLVING ENDOPEPTIDASES PEPO2
; FILE REFERENCE: WARF:0090US
; CURRENT APPLICATION NUMBER: US/10/873,427A
; CURRENT FILING DATE: 2004-06-21
; PRIOR APPLICATION NUMBER: 60/452,257
; PRIOR FILING DATE: 2003-06-20
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 643
; TYPE: PRT
; ORGANISM: Lactobacillus helveticus
US-10-873-427A-4

Query Match 36.1%; Score 39; DB 6; Length 643;
Best Local Similarity 40.0%; Pred. No. 1.4e+02;
Matches 8; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
Qy 3 NFDSIRVDVNDVADLLQI 22
```

Db 90 DFDNRNADGADPIQADLHEI 109

Search completed: February 10, 2006, 23:26:00  
Job time : 0.953629 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:05:33 ; Search time 18.0948 Seconds  
(without alignments)  
607.053 Million cell updates/sec

Title: US-10-797-821-26

Perfect score: 130

Sequence: 1 PLDKRGLNPLIHNSLVREVDRE 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	130	100.0	25	2	AAR43695	Aar43695 GT subseq
2	130	100.0	25	2	AAW34157	Aaw34157 GTF antig
3	130	100.0	25	2	AAy43337	Aay43337 GTF antig
4	130	100.0	25	7	ADD93646	Add93646 Streptoco
5	130	100.0	25	9	ADX37269	Adx37269 Streptoco
6	130	100.0	27	2	AAW34162	Aaw34162 GTF antig
7	130	100.0	1590	7	ADD93657	Add93657 Streptoco
8	130	100.0	1590	9	ADX37280	Adx37280 Streptoco
9	118	90.8	1592	2	AAR32925	Aar32925 Glucosylt
10	80	61.5	1017	5	AAU79285	Aau79285 Streptoco
11	80	61.5	1375	5	AAU98028	Aau98028 S. mutans
12	80	61.5	1375	5	AAU79288	Aau79288 Streptoco
13	80	61.5	1375	7	ADD93655	Add93655 Streptoco
14	80	61.5	1375	9	ADX37278	Adx37278 Streptoco
15	80	61.5	1475	5	AAU98031	Aau98031 S. mutans
16	80	61.5	1475	5	AAU98036	Aau98036 S. mutans
17	80	61.5	1475	5	AAU98037	Aau98037 S. mutans
18	80	61.5	1475	5	AAU98040	Aau98040 S. mutans
19	80	61.5	1475	5	AAU98035	Aau98035 S. mutans
20	80	61.5	1475	5	AAU98033	Aau98033 S. mutans
21	80	61.5	1475	5	AAU98034	Aau98034 S. mutans
22	80	61.5	1475	5	AAU98030	Aau98030 S. mutans
23	80	61.5	1475	5	AAU98039	Aau98039 S. mutans
24	80	61.5	1475	5	AAU98027	Aau98027 S. mutans

25	80	61.5	1475	5	AAU98032	Aau98032 S. mutans
26	80	61.5	1475	5	AAU98038	Aau98038 S. mutans
27	80	61.5	1475	7	ADD93654	Add93654 Streptoco
28	80	61.5	1475	9	ADX37277	Adx37277 Streptoco
29	80	61.5	1475	5	AAU79284	Aau79284 Streptoco
30	64	49.2	1499	7	ADC54806	Adc54806 Protein S
31	59	45.4	223	6	ABR63230	AbR63230 Glucanuc
32	59	45.4	2835	5	ABB98574	Abb98574 Dextran s
33	59	45.4	2835	6	ABR55594	AbR55594 Amino aci
34	58	44.6	224	6	ABR63227	AbR63227 Glucanuc
35	58	44.6	1497	6	ABR63234	AbR63234 Glucanuc
36	54	41.5	500	8	ADN19758	Adn19758 Bacterial
37	53	40.8	1430	5	AAU98043	Aau98043 S. mutans
38	53	40.8	1430	5	AAU98044	Aau98044 S. mutans
39	53	40.8	1430	5	AAU98045	Aau98045 S. mutans
40	53	40.8	1430	5	AAU98042	Aau98042 S. mutans
41	53	40.8	1430	5	AAU98029	Aau98029 S. mutans
42	53	40.8	1430	5	AAU98041	Aau98041 S. mutans
43	53	40.8	1430	7	ADD93656	Add93656 Streptoco
44	53	40.8	1430	9	ADX37279	Adx37279 Streptoco
45	51	39.2	221	6	ABR63226	AbR63226 Glucanuc

ALIGNMENTS

RESULT 1  
AAR43695  
ID AAR43695 standard; peptide; 25 AA.  
XX  
AC AAR43695;  
XX  
DT 25-MAR-2003 (revised)  
DT 20-MAY-1994 (first entry)  
XX  
XX GT subsequence.  
XX  
KW GT; glucosyltransferase; vaccine; T-cell; B-cell; reaction;  
KW immunoreponse; peptidyl core matrix; dental caries; diptheria; tetanus;  
KW measles; polio.  
XX  
OS Synthetic.  
XX  
PN WO9322341-A1.  
XX  
PD 11-NOV-1993.  
XX  
PF 30-APR-1993; 93WO-US004094.  
XX  
PR 01-MAY-1992; 92US-00877295.  
XX  
(FORS-) FORSVTH DENTAL INFIRMARY FOR CHILDREN.  
XX  
PI Taubman MA, Smith DJ;  
XX  
XX WPI; 1993-368721/46.  
XX  
PT Synthetic glucosyl-transferase peptide(s) - provoke T- and B-cell  
PT reactions in mammals, and are effective against dental caries.  
XX  
PS Claim 3; Page 23; 38pp; English.  
XX  
CC The sequences (AAR43694-98) are subsequences from GT, they provoke T- and  
CC B-cell responses in mammals. The can be used to produce a vaccine  
CC comprising 2 of these sequences attached to a peptidyl core matrix. They  
CC are useful in producing T-cell responses to the GT enzyme causing a  
CC reduction of colonisation or accumulation of mutans streptococcal strains  
CC in mammals. The vaccines can be used in preventing dental caries.  
CC (Updated on 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 25 AA;  
100.0%; Score 130; DB 2; Length 25;

Query Match

```

Best Local Similarity 100.0%; Pred. No. 3.6e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLDKRSGNLPLIHNSLVDRVDDRE 25
Db |||||

RESULT 2
ID AAW34157 standard; peptide; 25 AA.
XX AAW34157;
XX 25-MAR-2003 (revised)
DT 18-FEB-1998 (first entry)
XX GTF antigenic peptide #2.
XX Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;
KW surface domain; glucan-binding domain; mutans streptococcal strain;
KW immune response; glucan-binding activity; dental caries prevention.
XX Streptococcus mutans.
XX US5686075-A.
XX 11-NOV-1997.
XX 30-APR-1993; 93US-00057162.
XX 01-MAY-1992; 92US-00877295.
XX (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
XX Smith DJ, Taubman MA;
XX WPI; 1997-558089/51.
XX Immunogenic compositions containing streptococcal glucosyltransferase
PT peptide(s) - used for provoking immune response to streptococcal
PT glucosyltransferase for preventing dental caries.
XX Claim 1; Col 13; l1pp; English.
XX AAW34156-W34160 represent immunogenic fragments of the Streptomyces
CC mutans glucosyltransferase (GTF) enzyme. AAW34157 and AAW34158 are from
CC the catalytic domain of GTF, while AAW34160 is from the GTF surface
CC domain. AAW34156 and AAW34159 are from the glucan-binding domain of GTF.
CC These sequences can all be used in the immunogenic composition of the
CC invention. The composition of the invention can alternatively comprise
CC one or more of these sequences linked to a lysine core matrix (see
CC AAW34161-W34165). A composition comprising one of these sequences can be
CC administered to a mammal to raise an immune response, in a method for
CC interfering with the enzymatic activity of streptococcal
CC glucosyltransferase in a mammal. The immune response results in reduction
CC of the colonisation or accumulation of mutans streptococcal strains in
CC the mammal. Compositions containing AAW34156 specifically interfere with
CC the glucan-binding activity of the streptococcal glucosyltransferase. The
CC peptides can also be used in vaccines for preventing dental caries in
CC mammals. (Updated on 25-MAR-2003 to correct PF field.)
XX Sequence 25 AA;
Query Match 100.0%; Score 130; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 3.6e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLDKRSGNLPLIHNSLVDRVDDRE 25
Db |||||

RESULT 4
ID ADD93646 standard; peptide; 25 AA.
XX ADD93646;
XX
```

---

```

RESULT 3
ID AAY43337 standard; peptide; 25 AA.
XX AAY43337;
XX 26-JAN-2000 (first entry)
DT GTF antigenic epitope.
DE GTF; antigenic epitope; glucosyltransferase; immunogenic epitope; teeth;
KW immunogenic composition; streptococcal glucan; antibody response;
KW cariogenic bacteria; mutans streptococci; colonisation; caries;
KW dental caries; immunisation; therapy.
XX Synthetic.
XX WO9952548-A2.
XX 21-OCT-1999.
XX 09-APR-1999; 99WO-US007828.
XX 10-APR-1998; 98US-0081315P.
XX (LEES/) LEES A.
PA (TAUB/) TAUBMAN M A.
PA (SMIT/) SMITH D J.
XX Lees A, Taubman MA, Smith DJ;
PI WPI; 1999-620289/53.
XX Immunogenic compositions for control of dental caries, based on
PT Streptococcus mutans components, particularly for vaccination of infants.
XX Claim 11; Page 44; 54pp; English.
XX This sequence represents a glucosyltransferase (GTF) immunogenic epitope,
CC and can be used in the immunogenic composition of the invention. The
CC immunogenic composition (A) comprises at least one streptococcal glucan
CC (SG) and at least one component (I) covalently coupled, (in)directly, to
CC SG. (A) elicits antibodies (Ab) to both SG and (I). (A), and related
CC compositions, are used to induce an antibody response to cariogenic
CC bacteria, collectively called "mutans streptococci" (i.e. any of
CC Streptococcus mutans, S. rattus, S. cricetus or S. sobrinus), especially
CC to prevent colonisation, to reduce numbers of bacteria or to reduce
CC incidence of caries, in humans or other animals. The composition can
CC therefore be used for the treatment or prevention of caries (particularly
CC dental caries). Abs raised against (A) can be used: (a) to detect SG, and
CC (b) for passive immunisation, e.g. applied directly to the surface of
CC teeth. SG is a poorly immunogenic type 2 T-cell independent antigen but
CC when coupled to (I) it produces a conjugate which induces significant
CC immunity to both components, resulting in immunological memory and long-
CC lasting antibody production against SG, in adults and children
XX Sequence 25 AA;
Query Match 100.0%; Score 130; DB 2; Length 25;
Best Local Similarity 100.0%; Pred. No. 3.6e-14;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLDKRSGNLPLIHNSLVDRVDDRE 25
Db |||||

RESULT 4
ID ADD93646 standard; peptide; 25 AA.
XX ADD93646;
XX
```

DT 29-JAN-2004 (first entry)  
 XX Streptococcus mutans glucosyltransferase-B catalytic domain peptide.  
 DE  
 XX  
 XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.  
 KW  
 XX Streptococcus mutans.  
 OS  
 XX WO2003075845-A2.  
 PN  
 XX 18-SEP-2003.  
 PD  
 XX  
 XX 07-MAR-2003; 2003WO-US006962.  
 PF  
 XX  
 XX 07-MAR-2002; 2002US-0363209P.  
 PR  
 XX 08-AUG-2002; 2002US-0402483P.  
 PR  
 XX (FORSYTH INST.  
 PA  
 XX Smith DJ, Taubman MA;  
 PI WPI; 2003-845091/78.  
 XX  
 XX Composition useful as vaccines for dental caries comprises a fragment of  
 CC a glucan binding protein-B binding to a major histocompatibility complex  
 CC class II protein.  
 CC  
 XX Claim 20; Page 17; 49pp; English.  
 PS  
 XX The present sequence is that of a catalytic domain peptide fragment of  
 CC Streptococcus mutans glucosyltransferase. The peptide can be used in  
 CC immunogenic compositions and subunit vaccines for dental caries. These  
 CC compositions comprise a major histocompatibility complex (MHC) class II  
 CC protein-binding peptide from Streptococcus mutans glucan binding protein  
 CC -B (GbpB) covalently linked with a (preferably catalytic domain) peptide  
 CC subunit of a streptococcal glucosyltransferase. The compositions are used  
 CC in a claimed method of eliciting production of an antibody in a mammal.  
 CC Diabetic or multi-epitopic polypeptides can be prepared synthetically or  
 CC by recombinant DNA technology. Antibodies raised against MHC class II  
 CC binding fragments of GbpB can be used in passive immunisation.  
 XX  
 SQ Sequence 25 AA;  
 Query Match 100.0%; Score 130; DB 7; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-14;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PLDKRSGLNPLIHNSLVDRVDRE 25  
 DB 1 PLDKRSGLNPLIHNSLVDRVDRE 25  
 RESULT 5  
 ADX37269  
 ID ADX37269 standard; peptide; 25 AA.  
 XX  
 AC ADX37269;  
 XX  
 XX 21-APR-2005 (first entry)  
 DT  
 XX Streptococcus mutant glucan binding protein B peptide #26.  
 DE  
 XX immunogenicity; immune stimulation; glucan binding protein-B;  
 KW microparticle; major histocompatibility complex; tooth disease.  
 KW  
 XX Streptococcus mutans.  
 OS  
 XX US2005031633-A1.  
 PN  
 XX 10-FEB-2005.  
 PD  
 XX  
 XX 09-MAR-2004; 2004US-00797821.  
 PF  
 XX

PR 13-APR-1998; 98US-0081550P.  
 PR 08-JAN-1999; 99US-0115142P.  
 PR 12-APR-1999; 99US-00290049.  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 PR 07-MAR-2003; 2003US-00383930.  
 XX (SMIT/) SMITH D J.  
 PA (TAUB/) TAUBMAN M A.  
 XX Smith DJ, Taubman MA;  
 PI WPI; 2005-151644/16.  
 DR  
 XX New composition comprising a fragment of a glucan binding protein-B (GbpB) that binds to MHC class II protein, and a biocompatible microparticle, useful for producing an antibody (claimed) for immunizing mammals against dental caries.  
 PT  
 PT Claim 11; SEQ ID NO 26; 73pp; English.  
 XX  
 XX The invention relates to a composition comprising a fragment of a glucan binding protein-B (GbpB) and a biocompatible microparticle, where the fragment binds to a major histocompatibility complex (MHC) class II protein. The composition is useful for producing an antibody for immunizing mammals against dental caries. This sequence corresponds to a Streptococcus mutans GbpB-derived peptide of the invention.  
 CC  
 XX Sequence 25 AA;  
 SQ  
 Query Match 100.0%; Score 130; DB 9; Length 25;  
 Best Local Similarity 100.0%; Pred. No. 3.6e-14;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PLDKRSGLNPLIHNSLVDRVDRE 25  
 DB 1 PLDKRSGLNPLIHNSLVDRVDRE 25  
 RESULT 6  
 AAW34162  
 ID AAW34162 standard; peptide; 27 AA.  
 XX  
 AC AAW34162;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 18-FEB-1998 (first entry)  
 XX  
 DE GTF antigenic peptide #2 linked to polylysine core.  
 XX  
 KW Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;  
 KW surface domain; glucan-binding domain; mutans streptococcal strain;  
 KW immune response; glucan-binding activity; dental caries prevention.  
 XX  
 OS Synthetic.  
 OS Streptococcus mutans.  
 XX  
 XX Key Location/Qualifiers  
 XX Peptide 1..25  
 FT /label= GTF antigenic peptide #2 (see AAW34157)  
 FT /note= "attached to the dendritic polylysine core via the  
 FT alpha-amino group of Lys(26); a second copy of the  
 FT antigenic 25-mer is linked to Lys(26) via the omega amino  
 FT group"  
 FT Modified-site 26  
 FT /note= "Lys(26) is linked to one copy of the antigenic  
 FT peptide through the alpha-amino group, and to a second  
 FT copy of the peptide (not shown) via the omega amino  
 FT group"  
 FT Modified-site 27  
 FT /note= "the alpha amino acid group of Lys(27) forms a  
 FT peptide linkage with the carboxyl amino group of Lys(26);  
 FT the omega amino group of Lys(27) forms a peptide bond  
 FT

FT with a second Lys residue analogous to Lys(26)"

XX US5686075-A.

PN 11-NOV-1997.

XX 30-APR-1993; 93US-00057162.

XX 01-MAY-1992; 92US-00877295.

XX (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.

XX Smith DJ, Taubman MA;

XX WPI; 1997-558089/51.

XX Immunogenic compositions containing streptococcal glucosyltransferase peptide(s) - used for provoking immune response to streptococcal glucosyltransferase for preventing dental caries.

XX Claim 12; Col; 11pp; English.

XX AAW34161-W34165 represent the Streptomyces mutans glucosyltransferase (GTF) enzyme immunogenic fragments shown in AAW34156-W34160 linked to a polylysine core. AAW34157 and AAW34158 are from the catalytic domain of GTF, while AAW34160 is from the GTF surface domain. AAW34156 and AAW34159 are from the glucan-binding domain of GTF. These sequences, and the immunogenic fragments shown in AAW34156-W34160 can all be used in the immunogenic composition of the invention. A composition comprising one of these sequences can be administered to a mammal to raise an immune response, in a method for interfering with the enzymatic activity of streptococcal glucosyltransferase in a mammal. The immune response results in reduction of the colonisation or accumulation of mutans streptococcal strains in the mammal. Compositions containing AAW34156 specifically interfere with the glucan-binding activity of the streptococcal glucosyltransferase. The peptides can also be used in vaccines for preventing dental caries in mammals. (Updated on 25-MAR-2003 to correct Pf field.)

XX Sequence 27 AA;

Query Match 100.0%; Score 130; DB 2; Length 27;

Best Local Similarity 100.0%; Pred. No. 4e-14;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNSLVDRVDRE 25  
|||||

Db 1 PLDKRSGLNPLIHNSLVDRVDRE 25  
|||||

RESULT 7

ADD93657

ID ADD93657 standard; protein; 1590 AA.

XX AC ADD93657;

XX DT 29-JAN-2004 (first entry)

XX DE Streptococcus sobrinus glucosyltransferase-I.

XX KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

XX OS Streptococcus sobrinus.

XX WO2003075845-A2.

XX 18-SEP-2003.

XX 07-MAR-2003; 2003WO-US006962.

XX 07-MAR-2002; 2002US-0363209P.

XX 08-AUG-2002; 2002US-0402483P.

PA (FORS-) FORSYTH INST.

XX Smith DJ, Taubman MA;

XX WPI; 2003-845091/78.

XX Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.

XX Claim 16; Page 14; 49pp; English.

XX The present sequence is the protein sequence of Streptococcus sobrinus glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially from the catalytic domain of the polypeptide, can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from S. mutans glucan binding protein-B (GbpB) covalently linked with a peptide fragment of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diepitopic or multiepitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.

XX Sequence 1590 AA;

Query Match 100.0%; Score 130; DB 7; Length 1590;

Best Local Similarity 100.0%; Pred. No. 7.2e-12;

Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNSLVDRVDRE 25  
|||||

Db 520 PLDKRSGLNPLIHNSLVDRVDRE 544  
|||||

RESULT 8

ADX37280

ID ADX37280 standard; protein; 1590 AA.

XX AC ADX37280;

XX DT 21-APR-2005 (first entry)

XX DE Streptococcus sobrinus glucan binding protein B.

XX KW immunogenicity; immune stimulation; glucan binding protein-B;

XX KW microparticle; major histocompatibility complex; tooth disease.

XX OS Streptococcus sobrinus.

XX US2005031633-A1.

XX PD 10-FEB-2005.

XX PF 09-MAR-2004; 2004US-00797821.

XX PR 13-APR-1998; 98US-0081550P.

XX PR 08-JAN-1999; 99US-0115142P.

XX PR 12-APR-1999; 99US-00290049.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX PR 07-MAR-2003; 2003US-00383930.

XX (SMIT/) SMITH D J.

PA (TAUB/) TAUBMAN M A.

XX Smith DJ, Taubman MA;

XX WPI; 2005-151644/16.

XX New composition comprising a fragment of a glucan binding protein-B (GbpB) that binds to MHC class II protein, and a biocompatible



PT microparticle, useful for producing an antibody (claimed) for immunizing  
 PT mammals against dental caries.

PS Claim 7; SEQ ID NO 37; 73pp; English.

CC The invention relates to a composition comprising a fragment of a glucan  
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II  
 CC protein. The composition is useful for producing an antibody for  
 CC immunizing mammals against dental caries. This sequence corresponds to a  
 CC Streptococcus sobrinus GbpB protein of the invention.

XX Sequence 1590 AA;

Query Match 100.0%; Score 130; DB 9; Length 1590;  
 Best Local Similarity 100.0%; Pred. No. 7.2e-12;  
 Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNSLVDRVDDRE 25  
 DB 520 PLDKRSGLNPLIHNSLVDRVDDRE 544

RESULT 9

AAR32925  
 ID AAR32925 standard; protein; 1592 AA.

XX AAR32925;

DT 28-JUN-1993 (first entry)

DE Glucosyltransferase I.

XX GT-1; Streptococcus; dental; caries.

XX Streptococcus sobrinus.

XX JP05023188-A.

XX 02-FEB-1993.

XX 25-JUL-1991; 91JP-00186592.

XX 25-JUL-1991; 91JP-00186592.

XX (KATO/) KATO K.

XX (FUKU/) FUKUI I.

XX WPI; 1993-079449/10.

XX N-PSDB; AAQ37760.

PT DNA sequence glucosyl:transferase-I - comprises Streptococcus sobrinus  
 PT DNA sequence with at least one nucleotide added or deleted.

XX Claim 13; Page 15; 29pp; Japanese.

XX The DNA sequence from Streptococcus sobrinus strain 6715 encodes  
 CC glucosyltransferase-I (and mutants). The DNA was obtd. by treating S.  
 CC sobrinus 6715 with mutanolysin, extracting the chromosomal DNA, partially  
 CC digesting with Sau3AI and fractionating on agarose gel. The 3-5 kbp  
 CC fragment was ligated into pUC18 and E. coli JM109 transformed with it. A  
 CC GT-1 expressing clone was isolated and sequenced. The clone may be used  
 CC in the development of a drug for dental caries

XX Sequence 1592 AA;

Query Match 90.8%; Score 118; DB 2; Length 1592;  
 Best Local Similarity 92.0%; Pred. No. 7.2e-10;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNSLVDRVDDRE 25  
 DB 520 PTDVRSGLNPLIHNSLVDRVDDRE 544

RESULT 10

AAU79285

ID AAU79285 standard; protein; 1017 AA.

XX AAU79285;

DT 13-AUG-2002 (first entry)

DE Streptococcus mutans monoclonal antibody-related protein #2.

XX Antibody; dental caries; water insoluble glucan synthetase; anti-caries;  
 KW glucosyl transferase-B; immunotherapy.

OS Streptococcus mutans.

PN JP2002114709-A.

PD 16-APR-2002.

PF 04-OCT-2000; 2000JP-00304889.

PR 04-OCT-2000; 2000JP-00304889.

XX (UYN1-) UNIV NIPPON.

XX WPI; 2002-448885/48.

PT Anti-carries agent composed of a monoclonal antibody against an inhibitory  
 PT enzyme against water insoluble glucan synthetase of glucosyl transferase-  
 B (GTF-B) of Streptococcus mutans.

PS Claim 4; Page 17-19; 28pp; Japanese.

CC The invention relates to a monoclonal antibody against dental caries and  
 CC an anti-carries agent composed of a monoclonal antibody produced by  
 CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)  
 CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having  
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
 CC transferase-B. The monoclonal antibody specifically inhibits water  
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl  
 CC transferase-B and is used in the immunotherapy of dental caries. This  
 CC sequence represents a Streptococcus mutans monoclonal antibody-related  
 CC protein

XX Sequence 1017 AA;

Query Match 61.5%; Score 80; DB 5; Length 1017;  
 Best Local Similarity 64.0%; Pred. No. 0.00087;  
 Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNSLVDRVDDRE 25  
 DB 490 PLNQRSGMNPITNLSLVNRTDDNAE 514

RESULT 11

AAU98028

ID AAU98028 standard; protein; 1375 AA.

XX AAU98028;

DT 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFC.

XX Glucosyltransferase; GTFC; transgenic plant; paper sizing;

KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture.

XX Streptococcus mutans.

XX

```
PN US2002031826-A1.
XX
XX 14-MAR-2002.
XX
XX 19-DEC-2000; 2000US-00740274.
XX
XX 07-JUN-1995; 95US-00478704.
XX
XX 07-JUN-1995; 95US-00482711.
XX
XX 07-JUN-1995; 95US-00485243.
XX
XX 16-JAN-1998; 98US-00007999.
XX
XX 16-JAN-1998; 98US-00008172.
XX
XX 20-JAN-1998; 98US-00009620.
XX
XX 11-DEC-1998; 98US-00210361.
XX
XX (NICH/) NICHOLS S E.
XX
XX Nichols SE;
XX
XX WPI; 2002-414332/44.
XX
XX N-PSDB; ABK52939.
XX
XX Glucosyltransferase B or D protein useful for producing a glucan useful
XX as substitutes for and additions to modified starch and latexes in paper
XX manufacture, comprises mutations in specific positions.
XX
XX Disclosure; Page 30-33; 44pp; English.
XX
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
XX B polypeptide having changes at position from I448V, D457N, D567T,
XX K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,
XX I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a
XX GTF D polypeptide having changes at positions from T589D, T589E, N471D,
XX N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
XX GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
XX complementary polynucleotide, a ribonucleic acid sequence encoding the
XX GTF mutant, an expression cassette comprising the polynucleotide operably
XX linked to a promoter, a vector comprising the expression cassette, host
XX cell introduced with the vector, a transgenic plant comprising the
XX vector, a seed or tuber from the transgenic plant, a paper sining and/or
XX coating composition comprising a glucan produced in a plant transformed
XX with a gene encoding the mutant GTF, wild type or, starch, a latex,
XX thermoplastic molecule or their combinations or glucan and starch where
XX the glucan is produced in the amyloplast and/or vacuole or a maize line
XX deficient in starch biosynthesis, transformed with a gene encoding a
XX glucosyltransferase B or D enzyme, wild-type or mutant and a paper
XX comprising the glucan (paper sizing/coating agent). The vector is useful
XX for producing a glucan in a plant. The method comprises transforming a
XX plant cell with the vector, growing the plant cell under plant growing
XX conditions to produce a regenerated plant and inducing expression of the
XX polynucleotide for a time sufficient to produce the glucan in the
XX regenerated plant, where the vector contains a transit sequence from
XX ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
XX chlorophyll AB binding protein to produce a transgenic plant, and glucan
XX is produced in the amyloplast of potato or the vacuole of sugar beet.
XX Glucans are useful as substitutes for and additions to modified starch
XX and latexes in paper manufacture. Unlike prior art techniques, which
XX require input materials that produce chemical effluents, paper
XX manufacture utilising the glucan produced by GTF, which utilises
XX biologically produced input materials, is more cost-effective and
XX environmentally friendly. Moreover, glucans also exhibit thermoplastic
XX properties and impart gloss to the paper during coating step. The present
XX sequence represents GTFc
XX
XX Sequence 1375 AA;
XX
XX Query Match 61.5%; Score 80; DB 5; Length 1375;
XX Best Local Similarity 64.0%; Pred. No. 0.0013;
XX Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
XX
XX 1 PLDKRSGLNPLIHNSLVDRVDRE 25
XX ||::||::||::||::||::||::||::||::||::||:|
XX 550 PLNQRSGMNPITNSLVNRTDDNAE 574
XX
XX RESULT 13
XX ADD93655
XX ID ADD93655 standard; protein; 1375 AA.
XX
XX ADD93655;
XX
XX 29-JAN-2004 (first entry)
XX
XX Streptococcus mutans glucosyltransferase-C.
XX
XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
XX Streptococcus mutans.
XX
XX WO2003075845-A2.
XX
XX 19-SEP-2003.
XX
```

```
RESULT 12
AAU79288
ID AAU79288 standard; protein; 1375 AA.
XX
XX AAU79288;
XX
XX 13-AUG-2002 (first entry)
XX
XX Streptococcus mutans monoclonal antibody-related protein #5.
XX
XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;
XX glucosyl transferase-B; immunotherapy.
XX
XX Streptococcus mutans.
XX
XX JP2002114709-A.
XX
XX 16-APR-2002.
XX
XX 04-OCT-2000; 2000JP-00304889.
XX
XX 04-OCT-2000; 2000JP-00304889.
XX
XX (UYNI-) UNIV NIPPON.
XX
XX WPI; 2002-448885/48.
XX
XX Anti-carries agent composed of a monoclonal antibody against an inhibitory
XX enzyme against water insoluble glucan synthetase of glucosyl transferase-
XX B (GTF-B) of Streptococcus mutans.
XX
XX Disclosure; Page 22-25; 28pp; Japanese.
XX
XX The invention relates to a monoclonal antibody against dental caries and
XX an anti-carries agent composed of a monoclonal antibody produced by
XX Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)
XX or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having
XX inhibitive activity against water insoluble glucan synthetase of glucosyl
XX transferase-B. The monoclonal antibody specifically inhibits water
XX insoluble glucan synthetase of Streptococcus mutans produced glucosyl
XX transferase-B and is used in the immunotherapy of dental caries. This
XX sequence represents a Streptococcus mutans monoclonal antibody-related
XX protein
XX
XX Sequence 1375 AA;
XX
XX Query Match 61.5%; Score 80; DB 5; Length 1375;
XX Best Local Similarity 64.0%; Pred. No. 0.0013;
XX Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
XX
XX 1 PLDKRSGLNPLIHNSLVDRVDRE 25
XX ||::||::||::||::||::||::||::||::||::||:|
XX 550 PLNQRSGMNPITNSLVNRTDDNAE 574
XX
XX RESULT 13
XX ADD93655
XX ID ADD93655 standard; protein; 1375 AA.
XX
XX ADD93655;
XX
XX 29-JAN-2004 (first entry)
XX
XX Streptococcus mutans glucosyltransferase-C.
XX
XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
XX Streptococcus mutans.
XX
XX WO2003075845-A2.
XX
XX 19-SEP-2003.
XX
```

XX 07-MAR-2003; 2003WO-US006962.  
 PF 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 XX (FORS-) FORSYTH INST.  
 PA Smith DJ, Taubman MA;  
 PI WPI; 2003-845091/78.  
 DR  
 XX Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.  
 XX Claim 16; Page 13; 49pp; English.  
 PS  
 XX The present sequence is the protein sequence of Streptococcus mutans  
 CC Glucosyltransferase-C (GTF-C). Peptide fragments of GTF-C, especially  
 CC from the catalytic domain of the polypeptide, can be used in immunogenic  
 CC compositions and subunit vaccines for dental caries. These compositions  
 CC comprise a major histocompatibility complex (MHC) class II protein-  
 CC binding peptide from S. mutans glucan binding protein-B (GbpB)  
 CC covalently linked with a peptide fragment of a streptococcal  
 CC glucosyltransferase. The compositions are used in a claimed method of  
 CC eliciting production of an antibody in a mammal. Diepitopic or  
 CC multiepitopic polypeptides can be prepared synthetically or by  
 CC recombinant DNA technology. Antibodies raised against MHC class II  
 CC binding fragments of GbpB can be used in passive immunisation.  
 XX  
 SQ Sequence 1375 AA;  
 Query Match 61.5%; Score 80; DB 7; Length 1375;  
 Best Local Similarity 64.0%; Pred. No. 0.0013;  
 Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 PLDKRSGLNPLIHNSLVDRVDDRE 25  
 Db 550 PLNQRSGMPLITNSLVNRTDDNAE 574  
 RESULT 14  
 ADX37278  
 ID ADX37278 standard; protein; 1375 AA.  
 XX  
 AC ADX37278;  
 XX  
 DT 21-APR-2005 (first entry)  
 XX  
 DE Streptococcus mutant glucan binding protein B variant #7.  
 XX  
 KW immunogenicity; immune stimulation; glucan binding protein-B;  
 KW microparticle; major histocompatibility complex; tooth disease.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PN US2005031633-A1.  
 XX  
 PD 10-FEB-2005.  
 XX  
 PF 09-MAR-2004; 2004US-00797821.  
 XX  
 PR 13-APR-1998; 98US-0081550P.  
 PR 08-JAN-1999; 99US-0115142P.  
 PR 12-APR-1999; 99US-00290049.  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 PR 07-MAR-2003; 2003US-00383930.  
 XX  
 XX (SMIT/) SMITH D J.  
 PA (TAUB/) TAUBMAN M A.  
 XX

PI Smith DJ, Taubman MA;  
 XX WPI; 2005-151644/16.  
 DR  
 XX New composition comprising a fragment of a glucan binding protein-B  
 PT (GbpB) that binds to MHC class II protein, and a biocompatible  
 PT microparticle, useful for producing an antibody (claimed) for immunizing  
 PT mammals against dental caries.  
 XX  
 PS Claim 7; SEQ ID NO 35; 73pp; English.  
 XX  
 CC The invention relates to a composition comprising a fragment of a glucan  
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II  
 CC protein. The composition is useful for producing an antibody for  
 CC immunizing mammals against dental caries. This sequence corresponds to a  
 CC Streptococcus mutans GbpB protein of the invention.  
 XX  
 SQ Sequence 1375 AA;  
 Query Match 61.5%; Score 80; DB 9; Length 1375;  
 Best Local Similarity 64.0%; Pred. No. 0.0013;  
 Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 PLDKRSGLNPLIHNSLVDRVDDRE 25  
 Db 550 PLNQRSGMPLITNSLVNRTDDNAE 574  
 RESULT 15  
 AAU98031  
 ID AAU98031 standard; protein; 1475 AA.  
 XX  
 AC AAU98031;  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE S. mutans glucosyltransferase GTFB mutant D457N.  
 XX  
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.  
 XX  
 OS Streptococcus mutans.  
 OS Synthetic.  
 XX  
 PH Key Location/Qualifiers  
 FT Misc-difference 457  
 FT /note= "Wild-type Asp substituted by Asn"  
 XX  
 PN US2002031826-A1.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PF 19-DEC-2000; 2000US-00740274.  
 XX  
 PR 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 XX (NICH/) NICHOLS S E.  
 PA  
 XX Nichols SE;  
 PI  
 XX WPI; 2002-414332/44.  
 DR  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.

XX Claim 36; Page; 44pp; English.

PS The invention an isolated protein comprising a glucosyltransferase (GTF)

XX B polypeptide having changes at position from I448V, D457N, D567T, K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T, I448V/D457N/D567T/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a GTF D polypeptide having changes at positions from T589D, T589E, N471D, N471D/T589D, and N471D/T589E. Also included are a glucan produced by the GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its complementary polynucleotide, a ribonucleic acid sequence encoding the GTF mutant, an expression cassette comprising the polynucleotide operably linked to a promoter, a vector comprising the expression cassette, host cell introduced with the vector, a transgenic plant comprising the vector, a seed or tuber from the transgenic plant, a paper sizing and/or coating composition comprising a glucan produced in a plant transformed with a gene encoding the mutant GTF, wild type or, starch, a latex, thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents a GTFB mutant of the invention. Note: The present sequence is not shown in the specification but was created by the indexer using the GTFB sequence appearing as AAU98027 and the information in claim 36

XX SQ Sequence 1475 AA;

Query Match 61.5%; Score 80; DB 5; Length 1475;

Best Local Similarity 64.0%; Pred. No. 0.0014;

Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 PLDKRSGINPLIHNLSLVDRVDDRE 25

Db 524 PLNQRSGMPLIHNLSLVNRTDDNAE 548

|||||:|||||:|||||:|

Search completed: February 10, 2006, 22:19:43

Job time : 19.0948 secs



A;Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconostoc  
A;Reference number: A33135; MUID:87308013; PMID:3040685  
A;Accession: C33135  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-349 <SHI>  
A;Cross-references: UNIPARC:UPI000014E25D; GB:M17361  
C;Genetics:  
A;Gene: gtfC  
C;Function:

A;Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
A;Keywords: duplication; Glycosyltransferase; hexosyltransferase  
F;1-34/Domain: signal sequence #status predicted <SIG>  
F;35-1375/Product: glycosyltransferase #status predicted <MAT>  
F;1126-1145/Domain: cpl repeat homology <CP1>  
F;1253-1272/Domain: cpl repeat homology <CP2>  
F;1318-1337/Domain: cpl repeat homology <CP3>

Query Match 61.5%; Score 80; DB 2; Length 1375;  
Best Local Similarity 64.0%; Pred. No. 0.0011;  
Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNLSLVDRVDDRE 25  
||:||||:||||:||||:|  
Db 550 PLNQRSGMPLITNSLVNRTDDNAE 574

## RESULT 3

B33135  
gtfB protein precursor - Streptococcus mutans  
C;Species: Streptococcus mutans  
C;Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 09-Jul-2004  
C;Accession: B33135; A33128  
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A;Reference number: A33135; MUID:87308013; PMID:3040685

A;Accession: B33135  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1475 <SHI>  
A;Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:gl53639; PIDN:  
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
submitted to the Protein Sequence Database, September 1990

A;Reference number: A33128

A;Accession: A33128

A;Status: preliminary; not compared with conceptual translation

A;Molecule type: DNA  
A;Residues: 1-171,173-641,'N',643-1475 <SH2>  
A;Cross-references: UNIPARC:UPI000017AC5E  
A;Experimental source: strain GS-5  
F;1096-1115/Domain: cpl repeat homology <CP1>  
F;1224-1243/Domain: cpl repeat homology <CP2>  
F;1289-1308/Domain: cpl repeat homology <CP3>  
F;1354-1373/Domain: cpl repeat homology <CP4>  
F;1419-1438/Domain: cpl repeat homology <CP5>

Query Match 61.5%; Score 80; DB 2; Length 1475;  
Best Local Similarity 64.0%; Pred. No. 0.0012;  
Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNLSLVDRVDDRE 25  
||:||||:||||:||||:|  
Db 524 PLNQRSGMPLITNSLVNRTDDNAE 548

## RESULT 4

JCS473  
dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides  
C;Species: Leuconostoc mesenteroides  
C;Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 09-Jul-2004  
C;Accession: JCS473  
R;Monchois, V.; Willemot, R.M.; Rемаud-Simeon, M.; Croux, C.; Monsan, P.  
Gene 182, 23-32, 1996

A;Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconostoc  
A;Reference number: JCS473; MUID:97136686; PMID:8982063  
A;Accession: JCS473  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1290 <MON>  
A;Cross-references: UNIPROT:Q48756; UNIPARC:UPI000017ABDF; GB:U38181  
C;Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose onto  
C;Genetics:  
A;Gene: ddaA

A;Keywords: glycosyltransferase; hexosyltransferase  
F;78-870/Domain: catalytic #status predicted <CAT>  
F;922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 53.8%; Score 70; DB 2; Length 1290;  
Best Local Similarity 60.0%; Pred. No. 0.035; 8; Indels 0; Gaps 0;  
Matches 15; Conservative 2; Mismatches 8;

QY 1 PLDKRSGLNPLIHNLSLVDRVDDRE 25  
||:||||:||||:||||:|  
Db 359 PNDKRSGLPFTREHSLVKRITDDKE 383

## RESULT 5

T30857  
glucosyltransferase - Streptococcus salivarius  
C;Species: Streptococcus salivarius  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T30857  
R;Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A;Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for prin  
A;Reference number: Z20909; MUID:95122197; PMID:7822030

A;Accession: T30857  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1449 <SIM>  
A;Cross-references: UNIPROT:Q55264; UNIPARC:UPI00000B166B; EMBL:L35495; NID:G662378; PID  
A;Genetics:  
A;Gene: gtfL

Query Match 46.2%; Score 60; DB 2; Length 1449;  
Best Local Similarity 63.2%; Pred. No. 1.3;  
Matches 12; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNLSLVDR 19  
||:||||:||||:||||  
Db 579 PIGNRSGVEPLISNSLND 597

## RESULT 6

T30552  
glucosyltransferase N - Streptococcus salivarius (fragment)  
C;Species: Streptococcus salivarius  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T30552

R;Jaffe, R.I.  
submitted to the EMBL Data Library, February 1998  
A;Description: Streptococcus salivarius V1477 gtfN.  
A;Reference number: Z20854  
A;Accession: T30552

A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1449 <JAF>  
A;Cross-references: UNIPROT:O68542; UNIPARC:UPI00000B10FD; EMBL:AF049609; NID:G29335545; I  
C;Genetics:  
A;Gene: gtfN

Query Match 46.2%; Score 60; DB 2; Length 1449;  
Best Local Similarity 63.2%; Pred. No. 1.3;  
Matches 12; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNLSLVDR 19

```
Db          579  PIGNRSGVEPLNSLND 597
          :  |||:  ||| ||| ||| |||
RESULT 7
T40666
acetylglutamate synthase - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 15-Mar-2004
C;Accession: T40666
R;Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, November 1998
A;Reference number: Z21889
A;Accession: T40666
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-500 <LYN>
A;Cross-references: UNIPARC:UPI000006B3C4; EMBL:AL034352; PIDN:CAA22186.1; GSPDB:GN00067
A;Experimental source: strain 972h-; cosmid c725
C;Genetics:
A;Gene: SPDB:SPBC725.14
A;Map position: 2
C;Superfamily: amino-acid acetyltransferase, fungal type

Query Match          41.5%; Score 54; DB 2; Length 500;
Best Local Similarity 60.0%; Pred. No. 3.2;
Matches 12; Conservative 1; Mismatches 3; Indels 4; Gaps 1;

Qy          2  LDKRSGNPLIHNSLVDREV 21
          |||  ||| ||| ||| ||| :
Db          294  LDK-----NPLIHNLTDRSI 309

RESULT 8
A45866
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans
C;Species: Streptococcus mutans
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 15-Mar-2004
C;Accession: A45866
R;Honda, O.; Kato, C.; Kuramitsu, H.K.
J. Gen. Microbiol. 136, 2099-2105, 1990
A;Title: Nucleotide sequence of the Streptococcus mutans gtfD gene encoding the glucosyl
A;Reference number: A45866; MUID:91100958; PMID:2148600
A;Accession: A45866
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1431 <HON>
A;Cross-references: UNIPARC:UPI000017ACSC; GB:M29296
C;Keywords: glycosyltransferase; hexosyltransferase
F;181-201/Domain: cpl repeat homology <CP1>
F;1127-1146/Domain: cpl repeat homology <CP2>
F;1192-1211/Domain: cpl repeat homology <CP3>
F;1257-1276/Domain: cpl repeat homology <CP4>
F;1277-1297/Domain: cpl repeat homology <CP5>
F;1321-1340/Domain: cpl repeat homology <CP6>
F;1341-1361/Domain: cpl repeat homology <CP6>
F;1385-1404/Domain: cpl repeat homology <CP7>

Query Match          40.8%; Score 53; DB 2; Length 1431;
Best Local Similarity 40.6%; Pred. No. 16;
Matches 13; Conservative 5; Mismatches 6; Indels 8; Gaps 1;

Qy          1  PLDK-----RSGNPLIHNSLVDREVDDR 24
          |||  ||| ||| ||| ||| :
Db          538  PLEKASKNKEIRSGLEFVITNSLNRSASEGK 569

RESULT 9
T47243
amino-acid N-acetyltransferase (EC 2.3.1.1) [imported] - Neurospora crassa
N;Alternate names: acetylglutamate synthase
C;Species: Neurospora crassa
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 15-Mar-2004
```

```
C;Accession: T47243
R;Yu, Y.G.; Turner, G.E.; Weiss, R.L.
Mol. Microbiol. 22, 545-554, 1996
A;Title: Acetylglutamate synthase from Neurospora crassa: structure and regulation of ex
A;Reference number: Z24419; MUID:97093974; PMID:8939437
A;Accession: T47243
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-712 <YUY>
A;Cross-references: UNIPARC:UPI000006B8D1; EMBL:L35484; NID:9530155; PIDN:AAC37502.1; PI:
C;Genetics:
A;Introns: 102/3
A;Gene: arg-14
A;Superfamily: amino-acid acetyltransferase, fungal type
C;Keywords: acyltransferase; coenzyme A

Query Match          40.0%; Score 52; DB 2; Length 712;
Best Local Similarity 58.8%; Pred. No. 9.8;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy          5  RSGNPLIHNSLVDREV 21
          :  ||| ||| ||| :
Db          465  RTWRNPLIHNLTD RPI 481

RESULT 10
F84462
transcription factor ZAP1 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 31-Dec-2004
C;Accession: F84462
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; i
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Unayam, L.; Tallon, L.
euss, D.; Nierman, W.C.; White, O.; Eissen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A;Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A;Reference number: A84420; MUID:20083487; PMID:10617197
A;Accession: F84462
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-487 <STO>
A;Cross-references: UNIPROT:Q9SI37; UNIPARC:UPI000000A13FD; GB:AE002093; NID:94585919; PI:
C;Genetics:
A;Gene: At2g04880
A;Map position: 2
C;Superfamily: DNA-binding protein WRKY1

Query Match          39.2%; Score 51; DB 2; Length 487;
Best Local Similarity 61.5%; Pred. No. 9;
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy          13  HNSLVDREVDDRE 25
          ||::: ||| ||| :
Db          373  HNNMLDSEVDKXE 385

RESULT 11
T50534
hypothetical protein [imported] - Acidianus ambivalens
C;Species: Acidianus ambivalens
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
C;Accession: T50534
R;Gomes, C.M.; Lemos, R.S.; Teixeira, M.; Kletzin, A.; Huber, H.; Stetter, K.O.; Schaefer
Biochim. Biophys. Acta 1411, 134-141, 1999
A;Title: The unusual iron sulfur composition of the Acidianus ambivalens succinate dehyd
A;Reference number: Z25105; PMID:10216159
A;Accession: T50534
A;Status: preliminary; translated from GB/EMBL/DBDJ
A;Molecule type: DNA
A;Residues: 1-90 <GOM>
A;Cross-references: UNIPROT:O73935; UNIPARC:UPI00000628DB; EMBL:AJ005961; PIDN:CAA06778.
A;Experimental source: isolate Le1 10
```





Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	130	100.0	1590	2	Q55263_9STRE	Q55263 streptococc
2	130	100.0	1590	2	Q59983_9STRE	Q59983 streptococc
3	118	100.0	1597	1	GTF1_STRDO	P11001 streptococc
4	118	90.8	1592	1	GTF2_STRDO	P27470 streptococc
5	80	61.5	1455	1	GTF3_STRDM	P13470 streptococc
6	80	61.5	1476	1	GTFB_STRDM	P09887 streptococc
7	70	53.8	1290	2	Q48756_LEJUME	Q48756 leuconostoc
8	66	50.8	1454	2	Q69A94_LEJUME	Q69A94 leuconostoc
9	64	49.2	1016	2	Q9LCJ7_LEJUME	Q9LCJ7 leuconostoc
10	64	46.2	1449	2	Q6B542_STRSL	Q6B542 streptococc
11	60	46.2	1449	2	Q55264_STRSL	Q55264 streptococc
12	59	45.4	2835	2	Q8G9Q2_LEJUME	Q8G9Q2 leuconostoc
13	58	44.6	1561	2	Q5SBM8_9LACO	Q5SBM8 lactobacill
14	55	42.3	999	2	Q6FWY7_CANGA	Q6FWY7 candida gla
15	54	41.5	500	2	Q94330_SCHPO	Q94330 schizosacch
16	53	40.8	212	2	Q4YNW0_PLABE	Q4YNW0 plasmodium
17	53	40.8	1099	2	GTFPSW4_ANOGA	Q7P8W4 anopheles g
18	53	40.8	1462	1	GTFD_STRDM	P49331 streptococc
19	52.5	40.4	472	2	Q8KF59_CHLTE	Q8KF59 chlorobium
20	52.5	40.4	751	2	Q52MV8_CHICK	Q52MV8 gallus gall
21	52	40.0	455	2	Q5TUM5_ANOGA	Q5TUM5 anopheles g
22	52	40.0	521	2	Q6CRG8_KLUJA	Q6CRG8 kluveromyce
23	52	40.0	712	2	Q12643_NEUCR	Q12643 neurospora
24	52	40.0	1577	2	Q54178_STGRN	Q54178 streptococc
25	51	39.2	487	2	Q50UK4_ENTHI	Q50UK4 entamoeba h
26	51	39.2	487	1	WRKY1_ARATH	Q9SIJ37 arabidopsis
27	51	39.2	521	2	Q7VBP5_PROWA	Q7VBP5 prochloroco
28	51	39.2	566	2	Q6F575_CANGA	Q6F575 candida gla
29	51	39.2	1772	2	Q5SBN3_LACRE	Q5SBN3 lactobacill
30	51	39.2	1781	2	Q5SBL9_LACRE	Q5SBL9 lactobacill
31	50.5	38.8	416	2	Q6MM39_BDBBA	Q6MM39 bdellovibri

```
DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtfi;
OS Streptococcus sobrinus
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OMZ176;
RX MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
RT Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; IHCX.
DR GO; GO:0047849; F:dextranase activity; IEA.
DR GO; GO:0016757; F:transferase activity; transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glucosyltransferase; Signal; Transferase.
FT SIGNAL 39 1590
FT CHAIN 39 1590 glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 100.0%; Score 130; DB 2; Length 1590;
Best Local Similarity 100.0%; Pred. No. 3e-10;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLDKRSLNPLIHNSLVDRVDRE 25
Db 520 PLDKRSLNPLIHNSLVDRVDRE 544

RESULT 3
GTF1_STRDO ID -GTF1_STRDO STANDARD; PRT; 1597 AA.
AC P11001;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfi;
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=MFE28; PubMed=3040686;
RX MEDLINE=87308014; PubMed=3040686;
RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
RT sobrinus MFE28.";
RL J. Bacteriol. 169:4271-4278(1987).
CC -I- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -I- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-

fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes both
CC water-soluble glucans (alpha 1,6-glucose). GTF-Si synthesizes both
CC forms of glucans.
CC -I- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -I- SIMILARITY: Contains 19 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
EMBL; M17391; AAC63063.1; -; Genomic DNA.
DR InterPro; IPR002479; Cell_wall_bd_put.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Dental caries; Glucosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38
FT CHAIN 39 1597 Glucosyltransferase-I.
FT REPEAT 1099 1132 A repeat.
FT REPEAT 1163 1213 AC repeat.
FT REPEAT 1227 1277 AC repeat.
FT REPEAT 1292 1342 AC repeat.
FT REPEAT 1352 1399 B repeat.
FT REPEAT 1406 1455 AC repeat.
FT REPEAT 1465 1512 B repeat.
FT REPEAT 1519 1568 AC repeat.
FT REPEAT 1582 1597 A repeat (incomplete).
FT REGION 39 1050 Catalytic (approximate).
FT REGION 1099 1597 1.25 A, 2 B and 5 AC repeats.
FT REGION 1099 1597 Glucan-binding (approximate).
SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 100.0%; Score 130; DB 1; Length 1597;
Best Local Similarity 100.0%; Pred. No. 3e-10;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLDKRSLNPLIHNSLVDRVDRE 25
Db 526 PLDKRSLNPLIHNSLVDRVDRE 550

RESULT 4
GTF2_STRDO ID -GTF2_STRDO STANDARD; PRT; 1592 AA.
AC P27470;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=6715 / Serotype G;
RX MEDLINE=91123227; PubMed=1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
RA Kagawa H.;
RT "Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL J. Bacteriol. 173:989-996(1991).
CC -I- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
```

CC -|- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-  
 CC fructose + (1,6-alpha-D-glucosyl) (n+1).  
 CC -|- SUBCELLULAR LOCATION: Secreted.  
 CC -|- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha  
 CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes  
 CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both  
 CC forms of glucans.  
 CC -|- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.  
 CC -|- SIMILARITY: Contains 16 cell wall binding repeats.  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use as long as its content is in no way modified and this statement is not  
 CC removed.

CC EMBL; D90213; BAA1241.1; -; Genomic\_DNA.  
 DR HSSP; P06653; 1GVN.  
 DR InterPro; IPR002479; Cell wall bd put.  
 DR Pfam; PF01473; CW\_binding\_1; 3.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.  
 FT SIGNAL 1 38 Potential.  
 FT CHAIN 39 1592 Glucosyltransferase-I.  
 FT REPEAT 1093 1142 1.  
 FT REPEAT 1158 1207 2.  
 FT REPEAT 1222 1272 3.  
 FT REPEAT 1287 1337 4.  
 FT REPEAT 1402 1451 5.  
 FT REPEAT 1514 1563 6.  
 FT REPEAT 1577 1592 7 (incomplete).  
 FT REGION 39 1044 Catalytic (approximate).  
 FT REGION 1093 1592 7 X tandem repeats.  
 FT REGION 1093 1592 Glucan-binding (approximate).  
 FT SEQUENCE 1592 AA; 176168 MW; 80A66D079351ECF CRC64;

Query Match 90.8%; Score 118; DB 1; Length 1592;  
 Best Local Similarity 92.0%; Pred. No. 2e-08;  
 Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 PLDVRSGLNPLHNSLVDRVDDRE 25  
 Db 520 PTDVRSGLNPLHNSLVDRVDDRE 544

## RESULT 5

GTFC\_STRMU  
 ID GTFC\_STRMU STANDARD; PRT; 1455 AA.  
 AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-MAY-2005 (Rel. 47, Last annotation update)  
 DE Glucosyltransferase-SI precursor (EC 2.4.1.15) (GTF-SI)  
 DE (Dextrantransferase) (Sucrose 6-glucosyltransferase).  
 GN Name=gtfc; OrderedLocNames=SMU.1005;  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OC NCBI\_TaxID=1309;  
 OK [1]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=GS-5;  
 RX MEDLINE=89137980; PubMed=2976010; DOI=10.1016/0378-1119(88)90382-4;  
 RA Ueda S., Shiroza T., Kuramitsu H.K.;  
 RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";  
 RL Gene 69:101-109(1988).  
 RN [2]  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,  
 RC MT4467 / Serotype e, and MT8148 / Serotype c;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,

RA Kimura S., Hamada S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans.";  
 RL FEMS Microbiol. Lett. 161:331-336(1998).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=UAI59 / ATCC 700610 / Serotype C;  
 RX MEDLINE=2295063; PubMed=12397186; DOI=10.1073/pnas.172501299;  
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,  
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,  
 RA Ferretti J.J.;  
 RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental  
 RT pathogen.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
 RN [4]  
 RP NUCLEOTIDE SEQUENCE OF 1-349.  
 RC STRAIN=GS-5;  
 RX MEDLINE=87308013; PubMed=3040685;  
 RA Shiroza T., Ueda S., Kuramitsu H.K.;  
 RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";  
 RL J. Bacteriol. 169:4263-4270(1987).  
 CC -|- FUNCTION: Production of extracellular glucans, that are thought to  
 CC play a key role in the development of the dental plaque because of  
 CC their ability to adhere to smooth surfaces and mediate the  
 CC aggregation of bacterial cells and food debris.  
 CC -|- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-  
 CC fructose + (1,6-alpha-D-glucosyl) (n+1).  
 CC -|- SUBCELLULAR LOCATION: Secreted.  
 CC -|- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha  
 CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes both  
 CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both  
 CC forms of glucans.  
 CC -|- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.  
 CC -|- SIMILARITY: Contains 5 cell wall binding repeats.  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use as long as its content is in no way modified and this statement is not  
 CC removed.  
 CC EMBL; M22054; AAA88592.1; -; Genomic\_DNA.  
 DR EMBL; D88652; BAA26102.1; -; Genomic\_DNA.  
 DR EMBL; D88655; BAA26106.1; -; Genomic\_DNA.  
 DR EMBL; D88658; BAA26110.1; -; Genomic\_DNA.  
 DR EMBL; D88661; BAA26114.1; -; Genomic\_DNA.  
 DR EMBL; D89978; BAA26120.1; -; Genomic\_DNA.  
 DR EMBL; AB014940; AAN58706.1; -; Genomic\_DNA.  
 DR EMBL; M17361; AAA88589.1; -; Genomic\_DNA.  
 DR PIR; JTO345; JTO345.  
 DR HSSP; P06653; 1H8G.  
 DR InterPro; IPR002479; Cell wall bd put.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 2.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Complete proteome; Dental caries; Glycosyltransferase; Repeat; Signal;  
 KW Transferase.  
 FT SIGNAL 1 34  
 FT CHAIN 35 1455 Glucosyltransferase-SI.  
 FT REPEAT 1126 1159 A repeat.  
 FT REPEAT 1169 1200 A repeat.  
 FT REPEAT 1227 1238 C repeat.  
 FT REPEAT 1253 1303 AC repeat.  
 FT REPEAT 1318 1330 A repeat (incomplete).  
 FT REGION 35 1050 Catalytic (approximate).  
 FT REGION 1126 1455 Glucan-binding (approximate).  
 FT REGION 1126 1455 2.4 A, 1 C and 1 AC repeats.  
 FT VARIANT 21 21 V -> I (in strain GS-5).  
 FT VARIANT 81 81 P -> L (in strain MT4239).  
 FT VARIANT 106 106 D -> V (in strain GS-5).  
 FT VARIANT 116 116 S -> A (in strain GS-5 and strain  
 FT MT4467).

```
FT VARIANT 126 126 A -> T (in strain GS-5).
FT VARIANT 150 151 SR -> PK (in strain GS-5, strain MT4239
FT VARIANT 256 256 and strain MT4467).
FT VARIANT 425 425 A -> V (in strain GS-5 and strain
FT VARIANT 519 519 MT4467).
FT VARIANT 538 538 R -> N (in strain MT4251).
FT VARIANT 545 545 Y -> D (in strain MT4245 and strain
FT VARIANT 597 597 MT4251).
FT VARIANT 600 600 R -> K (in strain MT4245 and strain
FT VARIANT 601 601 MT4251).
FT VARIANT 614 614 A -> T (in strain GS-5).
FT VARIANT 727 727 M -> T (in strain GS-5).
FT VARIANT 734 734 T -> I (in strain MT8148).
FT VARIANT 964 964 L -> F (in strain MT4239).
FT VARIANT 1113 1113 N -> Y (in strain MT4239).
FT VARIANT 1118 1118 A -> T (in strain MT4239).
FT VARIANT 1204 1204 I -> V (in strain GS-5, strain MT4239,
FT VARIANT 1208 1208 strain MT4467 and strain MT8148).
FT VARIANT 1292 1294 DGH -> NGY (in strain GS-5, strain MT4467
FT VARIANT 1305 1369 Missing (in strain MT4245).
FT VARIANT 1326 1326 I -> V (in strain MT8148).
FT VARIANT 1331 1331 T -> A (in strain GS-5, strain MT4239,
FT VARIANT 1377 1377 strain MT4467 and strain MT8148).
FT VARIANT 1398 1398 R -> K (in strain MT8148).
FT VARIANT 1424 1424 D -> N (in strain MT8148).
FT VARIANT 1439 1439 V -> I (in strain MT4239 and strain
FT VARIANT 1444 1444 MT8148).
FT CONFLICT 1337 1455 S -> P (in strain MT8148).
FT ORLYFKSGVQAKGELITERGRIKYYDPNSGNEVRNRYVR
FT TSSGNWYFNGDYGALIGWHVVEGRRVYDFENGVRVASHD
FT QRHWYDYRRDFORGSSAVRFRHSRNGFFDNFFRF ->
FT HASILSLMVFRLRESSLQSVKVSNTMLIPKMKFVIM
FT (in Ref. 1).
SQ SEQUENCE 1455 AA; 162966 MW; 3CB455A99A4FEC86 CRC64;
Query Match 61.5%; Score 80; DB 1; Length 1455;
Best Local Similarity 64.0%; Pred. No. 0.01;
Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
QY 1 PLDKRSGLNPLIHSNLSVDREYDDRE 25
Db 550 PLNORSQGNPLITNSLVNRTDDNAE 574
RESULT 6
GTFB_STRMU STANDARD; PRT; 1476 AA.
AC P08987; O69384; O69387; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfI; OrderedLocusNames=SMU.1004;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
"Sequence analysis of the gtfB gene from Streptococcus mutans.";
J. Bacteriol. 169:4263-4270(1987).
[2]
NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
[3]
NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.B., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -I- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -I- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -I- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -I- SIMILARITY: Contains 10 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; M17361; AAA88588.1; -; Genomic DNA.
CC EMBL; D88651; BAA26101.1; -; Genomic DNA.
CC EMBL; D88654; BAA26105.1; -; Genomic DNA.
CC EMBL; D88657; BAA26109.1; -; Genomic DNA.
CC EMBL; D88660; BAA26113.1; -; Genomic DNA.
CC EMBL; D89977; BAA26119.1; -; Genomic DNA.
CC EMBL; AE014940; AAN58705.1; -; Genomic DNA.
CC PIR; B33135; B33135.
CC HSP; P06653; IH8G.
CC InterPro; IPR002479; Cell_wall_bd_put.
CC InterPro; IPR003318; Glyco_hydro_70.
CC Pfam; PF01473; CW_binding_1; 4.
CC Pfam; PF02324; Glyco_hydro_70; 1.
KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
KW Transferrase.
FT SIGNAL 1 34 Potential.
FT CHAIN 35 1476 Glucosyltransferase-I.
FT REPEAT 1097 1130 A repeat.
FT REPEAT 1161 1210 1.
FT REPEAT 1225 1275 2.
FT REPEAT 1290 1340 3.
FT REPEAT 1355 1405 4.
FT REPEAT 1420 1470 5.
FT REGION 35 1051 Catalytic (approximate).
FT REGION 1097 1476 Glucan-binding (approximate).
FT REGION 1161 1470 5 X tandem repeats.
FT VARIANT 62 62 S -> T (in strain MT4239).
FT VARIANT 65 65 T -> I (in strain GS-5).
FT VARIANT 68 68 V -> A (in strain GS-5, strain MT4245,
FT strain MT4251, strain MT4467 and strain
```

```
FT VARIANT 78 78
FT VARIANT 86 86
FT VARIANT 89 89
FT VARIANT 168 168
FT VARIANT 276 276
FT VARIANT 399 399
FT VARIANT 474 474
FT VARIANT 512 512
FT VARIANT 519 519
FT VARIANT 701 701
FT VARIANT 708 708
FT VARIANT 938 938
FT VARIANT 952 952
FT VARIANT 963 964
FT VARIANT 968 970
FT VARIANT 1086 1086
FT VARIANT 1158 1158
FT VARIANT 1163 1163
FT VARIANT 1168 1168
FT VARIANT 1182 1182
FT VARIANT 1234 1234
FT VARIANT 1263 1263
FT VARIANT 1263 1263
FT VARIANT 1264 1264
FT VARIANT 1272 1272
FT VARIANT 1329 1329
FT VARIANT 1394 1394
FT VARIANT 1402 1402
FT VARIANT 1459 1459
FT CONFLICT 570 570
FT CONFLICT 800 817
FT CONFLICT 1310 1310
FT CONFLICT 1476 AA; 165847 MW; 9C6E09F731B4CBFC CRC64;
SQ SEQUENCE 61.5%; Score 80; DB 1; Length 1476;
Query Match Best Local Similarity 64.0%; Pred. No. 0.011;
Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;
Qy 1 PLDKRSGNLNPLIHNLSLVDRVDDRE 25
Db 524 PLNQRSGNPLIHNLSLVNRITDDNAE 548
:::|||||:|||||:|:|
RESULT 7
Q48756 LEUME PRELIMINARY; PRT; 1290 AA.
AC Q48756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Dextranucrase.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NRRL B1299;
RA Monchois V., Willemot R.M., Renaud-Simeon M., Croux C., Monsan P.;
MEDLINE=97136686; PubMed=8982063; DOI=10.1016/S0378-1119(96)00443-X;
RA Monchois V., Willemot R.M., Renaud-Simeon M., Croux C., Monsan P.;
"Cloning and sequencing of a gene coding for a novel dextranucrase
from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-
6) and alpha (1-3) linkages.";
Gene 182:23-32(1996).
[2]
RN NUCLEOTIDE SEQUENCE.
RC STRAIN=NRRL B1299;
RA Monchois V.M., Willemot R.M., Renaud-Simeon M., Croux C., Monsan P.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U38181; AAB40875.1; -; Genomic_DNA.
DR PIR; JC5473; JC5473.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 1.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1290 AA; 145590 MW; 3555C2E96B749FAA CRC64;
Query Match 53.8%; Score 70; DB 2; Length 1290;
Best Local Similarity 60.0%; Pred. No. 0.3;
Matches 15; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
Qy 1 PLDKRSGNLNPLIHNLSLVDRVDDRE 25
Db 359 PDKRSGLEPTREHSLVKRITDDKE 383
|:|||||:|||||:|:|
RESULT 8
Q69A94 LEUME PRELIMINARY; PRT; 1454 AA.
ID Q69A94 LEUME PRELIMINARY;
AC Q69A94;
DT 23-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Dextranucrase (EC 2.4.1.5).
GN Name=darP;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IBT-PQ;
RA Fernandez-Vazquez J.L., Lopez-Munguia A., Olvera C.;
RT "Molecular characterization of a dextranucrase gene from Leuconostoc
mesenteroides IBT-PQ isolated from pulque.";
RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY504865; AAS79426.1; -; Genomic DNA.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 1.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Glycosyltransferase; Transferase.
SQ SEQUENCE 1454 AA; 160569 MW; 2DC7342963B6FD87 CRC64;
Query Match 50.8%; Score 66; DB 2; Length 1454;
Best Local Similarity 51.7%; Pred. No. 1.4;
Matches 15; Conservative 4; Mismatches 6; Indels 4; Gaps 1;
Qy 1 PLD---KRSGLNPLIHNLSLVDRVDDRE 25
Db 547 PVDHYGNKRSGLPEFLTNLSLVNRITDDSRD 575
|:|||||:|||||:|:|
RESULT 9
Q9LCJ7 LEUME PRELIMINARY; PRT; 1016 AA.
ID Q9LCJ7 LEUME PRELIMINARY;
AC Q9LCJ7;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
```

DE Dextranucrase.  
 GN Name=dsrt;  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=NRRL B-512F;  
 RX MEDLINE=20169623; PubMed=10705445;  
 RA Funane K., Mizuno K., Takahara H., Kobayashi M.;  
 RT "Gene encoding a dextranucrase-like protein in Leuconostoc  
 RL mesenteroides NRRL B-512F."  
 RL Biosci. Biotechnol. Biochem. 64:29-38(2000).  
 DR EMBL; AB020020; BAA90527.1; -; Genomic\_DNA.  
 DR HSSP; P06278; IVUS.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1016 AA; 110343 MW; 8896EFDE13CCCB47 CRC64;

Query Match 49.2%; Score 64; DB 2; Length 1016;  
 Best Local Similarity 56.5%; Pred. No. 1.9;  
 Matches 13; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 3 DKRSGLNPLIHNSLVDRVDDRE 25  
 :||||| : : ||||| :  
 Db 598 NKRSGLPEFLTNSLVDRINDSTD 620

RESULT 10  
 O68542\_STRSL PRELIMINARY; PRT; 1449 AA.  
 ID O68542\_STRSL PRELIMINARY; PRT; 1449 AA.  
 AC O68542\_STRSL PRELIMINARY; PRT; 1449 AA.  
 DT 01-AUG-1998 (TrEMBLrel. 07, Created)  
 DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Glucosyltransferase N (Fragment).  
 GN Name=gtfn;  
 OS Streptococcus salivarius.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1304;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=VI477;  
 RA Jaffe R.I.;  
 DR Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF049609; AAC05156.1; -; Genomic\_DNA.  
 DR PIR; T30552; T30552.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 2.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 FT NON TER 1449 1449  
 SQ SEQUENCE 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;

Query Match 46.2%; Score 60; DB 2; Length 1449;  
 Best Local Similarity 63.2%; Pred. No. 11;  
 Matches 12; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNSLVDR 19  
 | : ||||| : ||||| :  
 Db 579 PIGNRSGLPEFLTNSLVDR 597

RESULT 11  
 Q55264\_STRSL PRELIMINARY; PRT; 1449 AA.  
 ID Q55264\_STRSL PRELIMINARY; PRT; 1449 AA.  
 AC Q55264\_STRSL PRELIMINARY; PRT; 1449 AA.  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Glucosyltransferase precursor.  
 GN Name=gtfl;  
 OS Streptococcus salivarius.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1304;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=95122197; PubMed=7822030;  
 RA Simpson C.L., Giffard P.M., Jacques N.A.;  
 RT "Streptococcus salivarius ATCC 25975 possesses at least two genes  
 RL coding for primer-independent glucosyltransferases."  
 RL Infect. Immun. 63:609-621(1995).  
 DR EMBL; L35495; AAC41412.1; -; Genomic\_DNA.  
 DR PIR; T30857; T30857.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 2.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Signal; Transferase.  
 FT SIGNAL 1 35 Potential.  
 FT CHAIN 36 1449 glucosyltransferase.  
 SQ SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;

Query Match 46.2%; Score 60; DB 2; Length 1449;  
 Best Local Similarity 63.2%; Pred. No. 11;  
 Matches 12; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNSLVDR 19  
 | : ||||| : ||||| :  
 Db 579 PIGNRSGLPEFLTNSLVDR 597

RESULT 12  
 O8GSQ2\_LEUME PRELIMINARY; PRT; 2835 AA.  
 ID O8GSQ2\_LEUME PRELIMINARY; PRT; 2835 AA.  
 AC O8GSQ2\_LEUME PRELIMINARY; PRT; 2835 AA.  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Dextranucrase (EC 2.4.1.5) (Fragment).  
 GN Name=dsrB;  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX MEDLINE=22231661; PubMed=12270834;  
 RX DOI=10.1128/JB.184.20-5753-5761.2002;  
 RA Bozonnet S., Dols-Lafargue M., Fabre E., Pizzut S., Renaud-Simeon M.,  
 RA Mondan P., Willemot R.M.;  
 RT "Molecular characterisation of DSR-E, an alpha-1,2 linkage  
 RT synthesising dextranucrase with two catalytic domains."  
 RL J. Bacteriol. 184:5753-5761(2002).  
 DR EMBL; AJ430204; CAD22883.1; -; Genomic\_DNA.  
 DR GO; GO:0047849; F:dextranucrase activity; IEA.  
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 3.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 2.  
 KW Glucosyltransferase; Transferase.  
 FT NON TER 1 1  
 SQ SEQUENCE 2835 AA; 313265 MW; D03262CDD735399D CRC64;

Query Match 45.4%; Score 59; DB 2; Length 2835;  
 Best Local Similarity 56.0%; Pred. No. 34;  
 Matches 14; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Qy 1 PLDKRSGLNPLHNSLVDREVDRE 25  
 ||:|||||:|||||  
 Db 600 PLAQSGLEPLINTSLVKGKDATE 624

RESULT 13  
 Q5SBM8\_9LACO PRELIMINARY; PRT; 1561 AA.  
 AC Q5SBM8;  
 DT 01-FEB-2005 (TrEMBLrel. 29, Created)  
 DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)  
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)  
 DE Glucanase (EC 2.4.1.5).  
 OS Lactobacillus parabuchneri.  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 OC Lactobacillus.  
 OX NCBI\_TaxID=152331;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=33;  
 RX PubMed=15528655; DOI=10.1099/mic.0.27321-0;  
 RA Kralj S., van Geel-Schutten G.H., Dondorff M.M.G., Kirsanovs S.,  
 van der Maarel M.J.E.C., Dijkhuizen L.;  
 RT "Glucan synthesis in the genus Lactobacillus: isolation and  
 RT characterization of glucanase genes, enzymes and glucan products  
 RT from six different strains.";  
 RL Microbiology 150:3681-3690(2004).  
 DR ENBL; AY697432; AAU08006.1; -; Genomic DNA.  
 DR GO; GO:0047849; F:dextranase activity; IEA.  
 DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR02479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW binding 1; 1.  
 DR DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Glycosyltransferase; Transferase.  
 SQ SEQUENCE 1561 AA; 172213 MW; F3701E1A6A8439C9 CRC64;

Query Match 44.6%; Score 58; DB 2; Length 1561;  
 Best Local Similarity 52.2%; Pred. No. 25;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

Qy 1 PLDKRSGLNPLHNSLVDREVD 23  
 ||:|||||:|||||  
 Db 575 PVSQSGLEPLTSLVNTGDD 597

RESULT 14  
 Q6FWY7\_CANGA PRELIMINARY; PRT; 999 AA.  
 AC Q6FWY7;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Candida glabrata strain CBS138 chromosome C complete sequence.  
 GN OrderedLocNames=CAGL0C0181g;  
 OS Candida glabrata (Yeast) (Torulopsis glabrata).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5478;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=ATCC 2001 / CBS 138;  
 RX PubMed=15229592; DOI=10.1038/nature02579;  
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
 Lafontaine I., de Montigny J., Marck C., Neuveglise C., Tallia E.,  
 Goffard N., Frangul N., Aigle M., Anthonard V., Babour A., Barbe V.,  
 Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,  
 Boisrame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,  
 Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
 Kérriste A., Koszul R., Lemaître M., Lesur I., Ma L., Muller H.,

RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;  
 RT "The genome sequence of *Schizosaccharomyces pombe*.";  
 RL Nature 415:871-880(2002).  
 DR ENBL; AL034352; CAA22186.1; -; Genomic\_DNA.  
 DR PIR; T40666; T40666.  
 DR GeneDB\_Spombe; SPBC725.14; -.  
 DR GO; GO:0003991; F:acetylglutamate kinase activity; IEA.  
 DR GO; GO:0008415; F:acyltransferase activity; IEA.  
 DR GO; GO:0004042; F:amino-acid N-acetyltransferase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006526; F:arginine biosynthesis; IEA.  
 DR InterPro; IPR011190; DUF619.  
 DR InterPro; IPR011190; NAG synth\_fungal.  
 DR Pfam; PF04768; DUF619; 1\_  
 DR PIRSF; PIRSF007892; NAGS\_fungal; 1.  
 DR Acyltransferase; Complete proteome; Transferase.  
 KW SEQUENCE 500 AA; 55777 MW; 97FD57788EF4E451 CRC64;

Query Match 41.5%; Score 54; DB 2; Length 500;  
 Best Local Similarity 60.0%; Pred. No. 27;  
 Matches 12; Conservative 1; Mismatches 3; Indels 4; Gaps 1;

QY 2 LDKRSGLNPLIHNSLVDREV 21  
 ||| ||||| ||| :  
 Db 294 LDK----NPLIHNVLTDRSI 309

Search completed: February 10, 2006, 22:31:14  
 Job time : 20.1623 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Biocolloration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:31:39 ; Search time 4.51109 Seconds  
(without alignments)  
458.180 Million cell updates/sec

Title: US-10-797-821-26

Perfect score: 130

Sequence: 1 PLDKRSGLNPLHNSLVDRVDRE 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

- 1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/iaa/PTCUS\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130	100.0	25	1	US-08-057-162B-2
2	80	61.5	1375	2	US-09-210-361-4
3	80	61.5	1375	2	US-09-740-274-4
4	80	61.5	1475	2	US-09-007-999-2
5	80	61.5	1475	2	US-09-210-361-2
6	80	61.5	1475	2	US-09-740-274-2
7	53	40.8	545	2	US-09-604-957-4
8	53	40.8	545	2	US-09-995-749A-10
9	53	40.8	1430	2	US-09-008-172-2
10	53	40.8	1430	2	US-09-210-361-6
11	53	40.8	1430	2	US-09-740-274-6
12	51	39.2	535	2	US-09-604-957-7
13	51	39.2	535	2	US-09-995-749A-13
14	51	39.2	1278	2	US-09-604-957-3
15	51	39.2	1781	2	US-09-995-749A-2
16	50	38.5	862	2	US-09-538-092-627
17	49	37.7	1575	2	US-10-144-198-43
18	49	37.7	1584	2	US-10-144-198-24
19	48	36.9	140	2	US-09-053-197A-34
20	48	36.9	140	2	US-09-085-761A-34
21	48	36.9	264	2	US-09-712-363-183
22	48	36.9	303	2	US-09-540-236-2675
23	46	35.4	591	2	US-09-248-796A-18374
24	46	35.4	665	2	US-09-248-796A-17182
25	45.5	35.0	750	2	US-09-248-796A-14886
26	45	34.6	272	2	US-09-107-532A-4656
27	45	34.6	433	2	US-09-949-016-8546

28	45	34.6	816	2	US-09-252-991A-30090	Sequence 30090, A
29	45	34.6	1896	2	US-09-949-016-9508	Sequence 9508, Ap
30	44.5	34.2	424	2	US-09-252-991A-18784	Sequence 18784, A
31	44	33.8	102	2	US-09-270-767-60550	Sequence 60550, A
32	44	33.8	278	2	US-09-248-796A-19983	Sequence 19983, A
33	44	33.8	346	2	US-09-270-767-45057	Sequence 45057, A
34	44	33.8	368	2	US-09-107-433-4034	Sequence 4034, Ap
35	44	33.8	373	2	US-09-270-767-42838	Sequence 42838, A
36	44	33.8	703	2	US-09-252-991A-18391	Sequence 18391, A
37	44	33.8	703	2	US-09-248-796A-14529	Sequence 14529, A
38	44	33.8	740	2	US-09-583-110-5308	Sequence 5308, Ap
39	44	33.8	1043	2	US-08-928-361B-30	Sequence 30, Appl
40	44	33.8	1693	2	US-09-560-385A-4	Sequence 4, Appli
41	44	33.8	1693	2	US-09-560-385A-8	Sequence 8, Appli
42	44	33.8	1713	2	US-08-600-982-24	Sequence 24, Appl
43	44	33.8	1713	2	US-09-560-385A-6	Sequence 6, Appli
44	44	33.8	1713	2	US-09-538-092-1359	Sequence 1359, Ap
45	44	33.8	1713	4	PCT-US94-10261A-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1  
US-08-057-162B-2  
; Sequence 2, Application US/08057162B  
; Patent No. 5686075  
; GENERAL INFORMATION:  
; APPLICANT: Taubman, Martin A.  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR DENTAL CARIES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/057,162B  
; FILING DATE: 30-APR-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/877,295  
; FILING DATE: 01-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wagner, Richard W.  
; REGISTRATION NUMBER: 34,480  
; REFERENCE/DOCKET NUMBER: FDC92-01A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 25 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-057-162B-2

Query Match 100.0%; Score 130; DB 1; Length 25;  
Best Local Similarity 100.0%; Pred. No. 2.9e-14;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLHNSLVDRVDRE 25  
|||||

Query Match 61.5%; Score 80; DB 2; Length 1475;



```
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match          40.8%; Score 53; DB 2; Length 1430;
Best Local Similarity 40.6%; Pred. No. 12;
Matches 13; Conservative 5; Mismatches 6; Indels 8; Gaps 1;

QY 1 PLDK-----RSGLNPLIHNSLVREVDNR 24
Db 538 PLEKASNKNEIRSGLEPVITNSLNRSABGK 569

RESULT 11
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match          40.8%; Score 53; DB 2; Length 1430;
Best Local Similarity 40.6%; Pred. No. 12;
Matches 13; Conservative 5; Mismatches 6; Indels 8; Gaps 1;

QY 1 PLDK-----RSGLNPLIHNSLVREVDNR 24
Db 538 PLEKASNKNEIRSGLEPVITNSLNRSABGK 569

US-09-604-957-7
; Sequence 7, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-604-957-7

Query Match          39.2%; Score 51; DB 2; Length 535;
Best Local Similarity 47.6%; Pred. No. 7.7;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 5 RSGLNPLIHNSLVREVDNR 25
Db 120 RWGLDAIVHQSLADRENNSTE 140

RESULT 13
US-09-995-749A-13
; Sequence 13, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 535
; TYPE: PRT
; ORGANISM: Lactobacillus reuteri
US-09-995-749A-13

Query Match          39.2%; Score 51; DB 2; Length 535;
Best Local Similarity 47.6%; Pred. No. 7.7;
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 5 RSGLNPLIHNSLVREVDNR 25
Db 120 RWGLDAIVHQSLADRENNSTE 140
```

RESULT 14  
 US-09-604-957-3  
 ; Sequence 3, Application US/09604957  
 ; Patent No. 6486314  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
 ; APPLICANT: DIJKHUIZEN, LUBBERT  
 ; APPLICANT: RAHAOUI, HAKIM  
 ; APPLICANT: LEER, ROBERT-JAN  
 ; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN  
 ; FILE REFERENCE: BO 43388  
 ; CURRENT APPLICATION NUMBER: US/09/604,957  
 ; CURRENT FILING DATE: 2000-06-28  
 ; PRIOR APPLICATION NUMBER: 00201871.1  
 ; PRIOR FILING DATE: 2000-05-25  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 3  
 ; LENGTH: 1278  
 ; TYPE: PRT  
 ; ORGANISM: Lactobacillus reuteri  
 US-09-604-957-3

Query Match 39.2%; Score 51; DB 2; Length 1278;  
 Best Local Similarity 47.6%; Pred. No. 22;  
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 Qy 5 RSGLNPLIHNSLVDREVDRE 25  
 |||: :||| :  
 Db 596 RWGLDAIVHQSLADRENNSTE 616

RESULT 15  
 US-09-995-749A-2  
 ; Sequence 2, Application US/09995749A  
 ; Patent No. 6867026  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
 ; APPLICANT: DIJKHUIZEN, LUBBERT  
 ; APPLICANT: RAHAOUI, HAKIM  
 ; APPLICANT: LEER, ROBERT-JAN  
 ; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
 ; FILE REFERENCE: B043388-CIP  
 ; CURRENT APPLICATION NUMBER: US/09/995,749A  
 ; CURRENT FILING DATE: 2001-11-29  
 ; PRIOR APPLICATION NUMBER: 09/604,957  
 ; PRIOR FILING DATE: 2000-06-28  
 ; PRIOR APPLICATION NUMBER: EPO 00201871.1  
 ; PRIOR FILING DATE: 2000-05-25  
 ; NUMBER OF SEQ ID NOS: 19  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 2  
 ; LENGTH: 1781  
 ; TYPE: PRT  
 ; ORGANISM: Lactobacillus reuteri  
 US-09-995-749A-2

Query Match 39.2%; Score 51; DB 2; Length 1781;  
 Best Local Similarity 47.6%; Pred. No. 33;  
 Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;  
 Qy 5 RSGLNPLIHNSLVDREVDRE 25  
 |||: :||| :  
 Db 1099 RWGLDAIVHQSLADRENNSTE 1119

Search completed: February 10, 2006, 22:36:15  
 Job time : 5.51109 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:14:04 ; Search time 14.6673 Seconds  
(without alignments)  
712.176 Million cell updates/sec

Title: US-10-797-821-26

Perfect score: 130

Sequence: 1 PLDKRSLNPLIHNLSLVREVDRE 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_Main:\*

1: /cgn2\_6/prodata/1/pubpaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pep:\*

3: /cgn2\_6/prodata/1/pubpaa/US09\_PUBCOMB.pep:\*

4: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pep:\*

5: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pep:\*

6: /cgn2\_6/prodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	130	100.0	25	4	US-10-383-930-26
2	130	100.0	25	5	US-10-797-821-26
3	130	100.0	1590	4	US-10-383-930-37
4	130	100.0	1590	5	US-10-797-821-37
5	80	61.5	1375	3	US-09-740-274-4
6	80	61.5	1375	4	US-10-383-930-35
7	80	61.5	1375	5	US-10-797-821-35
8	80	61.5	1475	3	US-09-740-274-2
9	80	61.5	1475	4	US-10-383-930-34
10	80	61.5	1475	5	US-10-797-821-34
11	59	45.4	223	5	US-10-484-218-10
12	58	44.6	224	5	US-10-484-218-4
13	58	44.6	1497	5	US-10-484-218-18
14	54	41.5	500	4	US-10-369-493-2411
15	53	40.8	545	3	US-09-995-749A-10
16	53	40.8	1430	3	US-09-740-274-6
17	53	40.8	1430	4	US-10-383-930-36
18	53	40.8	1430	5	US-10-797-821-36
19	51	39.2	221	5	US-10-484-218-2
20	51	39.2	265	4	US-10-425-114-37000
21	51	39.2	463	4	US-10-225-066A-924
22	51	39.2	463	4	US-10-374-780A-2242
23	51	39.2	463	5	US-10-225-066A-924
24	51	39.2	525	5	US-10-484-218-23
25	51	39.2	535	3	US-09-995-749A-13
26	51	39.2	1777	5	US-10-484-218-12
27	51	39.2	1781	3	US-09-995-749A-2

28	50	38.5	108	4	US-10-424-599-146923	Sequence 146923,
29	50	38.5	628	4	US-10-437-963-174471	Sequence 174471,
30	50	38.5	697	4	US-10-183-687-244	Sequence 244, App
31	49	37.7	695	4	US-10-108-260A-4190	Sequence 4190, App
32	49	37.7	695	4	US-10-648-593-168	Sequence 168, App
33	49	37.7	758	4	US-10-437-963-133898	Sequence 133898,
34	49	37.7	758	4	US-10-437-963-133896	Sequence 133896,
35	49	37.7	833	4	US-10-250-615-21	Sequence 21, Appl
36	49	37.7	1394	3	US-09-864-761-36118	Sequence 36118, A
37	49	37.7	1575	4	US-10-144-198-43	Sequence 24, Appl
38	49	37.7	1584	4	US-10-144-198-24	Sequence 15, Appl
39	49	37.7	1584	5	US-10-494-940-16	Sequence 34, Appl
40	48	36.9	140	4	US-10-023-171-34	Sequence 34, Appl
41	48	36.9	284	3	US-09-712-363-183	Sequence 183, App
42	48	36.9	264	4	US-10-282-122A-62339	Sequence 62339, A
43	48	36.9	264	4	US-10-282-122A-64483	Sequence 64483, A
44	48	36.9	512	6	US-11-097-143-23826	Sequence 23826, A
45	47	36.2	187	4	US-10-767-701-62530	Sequence 62530, A

ALIGNMENTS

RESULT 1

US-10-383-930-26

; Sequence 26, Application US/10383930

; Publication No. US20040127400A1

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J

; APPLICANT: Taubman, Martin A

; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein

; FILE REFERENCE: 25669-018

; CURRENT APPLICATION NUMBER: US/10/383,930

; CURRENT FILING DATE: 2003-03-07

; PRIOR APPLICATION NUMBER: 60/402,483

; PRIOR FILING DATE: 2002-08-08

; PRIOR APPLICATION NUMBER: 60/363,209

; PRIOR FILING DATE: 2002-03-07

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 25

; LENGTH: 25

; TYPE: PRT

; ORGANISM: Streptococcus mutans

US-10-383-930-26

Query Match 100.0%; Score 130; DB 4; Length 25;  
Best Local Similarity 100.0%; Pred. No. 38-12;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PLDKRSLNPLIHNLSLVREVDRE 25  
|||||  
Db 1 PLDKRSLNPLIHNLSLVREVDRE 25

RESULT 2

US-10-797-821-26

; Sequence 26, Application US/10797821

; Publication No. US20050031633A1

; GENERAL INFORMATION:

; APPLICANT: Smith, Daniel J.

; APPLICANT: Taubman, Martin A.

; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens

; FILE REFERENCE: 25669-020

; CURRENT APPLICATION NUMBER: US/10/797,821

; CURRENT FILING DATE: 2004-03-09

; PRIOR APPLICATION NUMBER: 10/383,930

; PRIOR FILING DATE: 2003-03-07

; PRIOR APPLICATION NUMBER: 60/363,209

; PRIOR FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: 60/402,483

; PRIOR FILING DATE: 2002-08-08

; PRIOR APPLICATION NUMBER: 09/290,049

; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 26  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Catalytic Domain GTF peptide  
US-10-797-821-26

Query Match 100.0%; Score 130; DB 5; Length 25;  
Best Local Similarity 100.0%; Pred. No. 3e-12;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNSLVDRVDDRE 25  
|||  
Db 1 PLDKRSGLNPLIHNSLVDRVDDRE 25

RESULT 3  
US-10-383-930-37  
; Sequence 37, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 37  
; LENGTH: 1590  
; TYPE: PRT  
; ORGANISM: Streptococcus sobrinus  
US-10-383-930-37

Query Match 100.0%; Score 130; DB 4; Length 1590;  
Best Local Similarity 100.0%; Pred. No. 3.3e-10;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNSLVDRVDDRE 25  
|||  
Db 520 PLDKRSGLNPLIHNSLVDRVDDRE 544

RESULT 4  
US-10-797-821-37  
; Sequence 37, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; PRIOR FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08

; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 37  
; LENGTH: 1590  
; TYPE: PRT  
; ORGANISM: Streptococcus sobrinus  
US-10-797-821-37

Query Match 100.0%; Score 130; DB 5; Length 1590;  
Best Local Similarity 100.0%; Pred. No. 3.3e-10;  
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNSLVDRVDDRE 25  
|||  
Db 520 PLDKRSGLNPLIHNSLVDRVDDRE 544

RESULT 5  
US-09-740-274-4  
; Sequence 4, Application US/09740274  
; Patent No. US20020031826A1  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 61.5%; Score 80; DB 3; Length 1375;  
Best Local Similarity 64.0%; Pred. No. 0.0086;  
Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNSLVDRVDDRE 25  
|||  
Db 550 PLNQRSGMNPITNSLVNRTDDNAE 574

RESULT 6  
US-10-383-930-35  
; Sequence 35, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018



; CURRENT APPLICATION NUMBER: US/10/383,930  
 ; CURRENT FILING DATE: 2003-03-07  
 ; PRIOR APPLICATION NUMBER: 60/402,483  
 ; PRIOR FILING DATE: 2002-08-08  
 ; PRIOR APPLICATION NUMBER: 60/363,209  
 ; PRIOR FILING DATE: 2002-03-07  
 ; NUMBER OF SEQ ID NOS: 41  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 35  
 ; LENGTH: 1375  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus mutans  
 US-10-383-930-35

Query Match 61.5%; Score 80; DB 4; Length 1375;  
 Best Local Similarity 64.0%; Pred. No. 0.0086;  
 Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PLDKRSGNPLIHNLSLVDRVDDRE 25  
 ||:||||:||||:||||:|  
 Db 550 PLNQRSGMPLITNSLVNRTDDNAE 574

RESULT 7  
 US-10-797-821-35  
 ; Sequence 35, Application US/10797821  
 ; Publication No. US20050031633A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Smith, Daniel J.  
 ; APPLICANT: Taubman, Martin A.  
 ; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
 ; FILE REFERENCE: 25669-020  
 ; CURRENT APPLICATION NUMBER: US/10/797,821  
 ; CURRENT FILING DATE: 2004-03-09  
 ; PRIOR APPLICATION NUMBER: 10/383,930  
 ; PRIOR FILING DATE: 2003-03-07  
 ; PRIOR APPLICATION NUMBER: 60/363,209  
 ; PRIOR FILING DATE: 2002-03-07  
 ; PRIOR APPLICATION NUMBER: 60/402,483  
 ; PRIOR FILING DATE: 2002-08-08  
 ; PRIOR APPLICATION NUMBER: 09/290,049  
 ; PRIOR FILING DATE: 1999-04-12  
 ; PRIOR APPLICATION NUMBER: 60/081,550  
 ; PRIOR FILING DATE: 1998-04-13  
 ; PRIOR APPLICATION NUMBER: 60/115,142  
 ; PRIOR FILING DATE: 1999-01-08  
 ; NUMBER OF SEQ ID NOS: 45  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 35  
 ; LENGTH: 1375  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus mutans  
 US-10-797-821-35

Query Match 61.5%; Score 80; DB 5; Length 1375;  
 Best Local Similarity 64.0%; Pred. No. 0.0086;  
 Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PLDKRSGNPLIHNLSLVDRVDDRE 25  
 ||:||||:||||:||||:|  
 Db 550 PLNQRSGMPLITNSLVNRTDDNAE 574

RESULT 8  
 US-09-740-274-2  
 ; Sequence 2, Application US/09740274  
 ; Patent No. US20020031826A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nichols, Scott E.  
 ; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
 ; FILE REFERENCE: 0357CRD  
 ; CURRENT APPLICATION NUMBER: US/09/740,274  
 ; CURRENT FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: 09/210,361  
 ; PRIOR FILING DATE: 1998-12-11  
 ; PRIOR APPLICATION NUMBER: 09/007,999  
 ; PRIOR FILING DATE: 1998-01-16  
 ; PRIOR APPLICATION NUMBER: 08/478,704  
 ; PRIOR FILING DATE: 1995-06-07  
 ; PRIOR APPLICATION NUMBER: 09/009,620  
 ; PRIOR FILING DATE: 1998-01-20  
 ; PRIOR APPLICATION NUMBER: 08/485,243  
 ; PRIOR FILING DATE: 1995-06-07  
 ; PRIOR APPLICATION NUMBER: 09/008,172  
 ; PRIOR FILING DATE: 1998-01-16  
 ; PRIOR APPLICATION NUMBER: 08/482,711  
 ; PRIOR FILING DATE: 1995-06-07  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 1475  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus mutans  
 US-09-740-274-2

Query Match 61.5%; Score 80; DB 3; Length 1475;  
 Best Local Similarity 64.0%; Pred. No. 0.0093;  
 Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PLDKRSGNPLIHNLSLVDRVDDRE 25  
 ||:||||:||||:||||:|  
 Db 524 PLNQRSGMPLITNSLVNRTDDNAE 548

RESULT 9  
 US-10-383-930-34  
 ; Sequence 34, Application US/10383930  
 ; Publication No. US20040127400A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Smith, Daniel J.  
 ; APPLICANT: Taubman, Martin A.  
 ; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
 ; FILE REFERENCE: 25669-018  
 ; CURRENT APPLICATION NUMBER: US/10/383,930  
 ; CURRENT FILING DATE: 2003-03-07  
 ; PRIOR APPLICATION NUMBER: 60/402,483  
 ; PRIOR FILING DATE: 2002-08-08  
 ; PRIOR APPLICATION NUMBER: 60/363,209  
 ; PRIOR FILING DATE: 2002-03-07  
 ; NUMBER OF SEQ ID NOS: 41  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 34  
 ; LENGTH: 1475  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus mutans  
 US-10-383-930-34

Query Match 61.5%; Score 80; DB 4; Length 1475;  
 Best Local Similarity 64.0%; Pred. No. 0.0093;  
 Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PLDKRSGNPLIHNLSLVDRVDDRE 25  
 ||:||||:||||:||||:|  
 Db 524 PLNQRSGMPLITNSLVNRTDDNAE 548

RESULT 10  
 US-10-797-821-34  
 ; Sequence 34, Application US/10797821  
 ; Publication No. US20050031633A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Smith, Daniel J.  
 ; APPLICANT: Taubman, Martin A.  
 ; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
 ; FILE REFERENCE: 25669-020  
 ; CURRENT APPLICATION NUMBER: US/10/797,821

; CURRENT FILING DATE: 2004-03-09  
 ; PRIOR APPLICATION NUMBER: 10/383,930  
 ; PRIOR FILING DATE: 2003-03-07  
 ; PRIOR APPLICATION NUMBER: 60/363,209  
 ; PRIOR FILING DATE: 2002-03-07  
 ; PRIOR APPLICATION NUMBER: 60/402,483  
 ; PRIOR FILING DATE: 2002-08-08  
 ; PRIOR APPLICATION NUMBER: 09/290,049  
 ; PRIOR FILING DATE: 1999-04-12  
 ; PRIOR APPLICATION NUMBER: 60/081,550  
 ; PRIOR FILING DATE: 1998-04-13  
 ; PRIOR APPLICATION NUMBER: 60/115,142  
 ; PRIOR FILING DATE: 1999-01-08  
 ; NUMBER OF SEQ ID NOS: 45  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 34  
 ; LENGTH: 1475  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus mutans  
 US-10-797-821-34

Query Match 61.5%; Score 80; DB 5; Length 1475;  
 Best Local Similarity 64.0%; Pred. No. 0.0093;  
 Matches 16; Conservative 5; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 PLDKRSGLNPLIHNSLVDRVDDRE 25  
 Db 524 PLNQRSGMPLITNSLVNRTDDNAE 548

RESULT 11  
 US-10-484-218-10  
 ; Sequence 10, Application US/10484218  
 ; Publication No. US20050059633A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
 ; TITLE OF INVENTION: GLUCANS AND GLUCANSUCRASES DERIVED FROM  
 ; TITLE OF INVENTION: LACTIC ACID BACTERIA  
 ; FILE REFERENCE: 2001-1316  
 ; CURRENT APPLICATION NUMBER: US/10/484,218  
 ; CURRENT FILING DATE: 2004-01-20  
 ; PRIOR APPLICATION NUMBER: PCT/NL02/00495  
 ; PRIOR FILING DATE: 2002-07-22  
 ; PRIOR APPLICATION NUMBER: EP 01202752.0  
 ; PRIOR FILING DATE: 2001-07-20  
 ; PRIOR APPLICATION NUMBER: EP 01202841.1  
 ; PRIOR FILING DATE: 2001-07-25  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: PatentIn Ver. 3.2  
 ; SEQ ID NO 10  
 ; LENGTH: 223  
 ; TYPE: PRT  
 ; ORGANISM: Lactobacillus sp.  
 US-10-484-218-10

Query Match 45.4%; Score 59; DB 5; Length 223;  
 Best Local Similarity 56.0%; Pred. No. 1.5;  
 Matches 14; Conservative 2; Mismatches 9; Indels 0; Gaps 0;  
 QY 1 PLDKRSGLNPLIHNSLVDRVDDRE 25  
 Db 112 PLAAQSGLEPLINTSLVKRCKDATE 136

RESULT 12  
 US-10-484-218-4  
 ; Sequence 4, Application US/10484218  
 ; Publication No. US20050059633A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
 ; TITLE OF INVENTION: GLUCANS AND GLUCANSUCRASES DERIVED FROM  
 ; TITLE OF INVENTION: LACTIC ACID BACTERIA  
 ; FILE REFERENCE: 2001-1316

; CURRENT APPLICATION NUMBER: US/10/484,218  
 ; CURRENT FILING DATE: 2004-01-20  
 ; PRIOR APPLICATION NUMBER: PCT/NL02/00495  
 ; PRIOR FILING DATE: 2002-07-22  
 ; PRIOR APPLICATION NUMBER: EP 01202752.0  
 ; PRIOR FILING DATE: 2001-07-20  
 ; PRIOR APPLICATION NUMBER: EP 01202841.1  
 ; PRIOR FILING DATE: 2001-07-25  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: PatentIn Ver. 3.2  
 ; SEQ ID NO 4  
 ; LENGTH: 224  
 ; TYPE: PRT  
 ; ORGANISM: Lactobacillus sp.  
 US-10-484-218-4

Query Match 44.6%; Score 58; DB 5; Length 224;  
 Best Local Similarity 52.2%; Pred. No. 2.2;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNSLVDRVDD 23  
 Db 112 PVSQSGLEPELTTSLVNRTGDD 134

RESULT 13  
 US-10-484-218-18  
 ; Sequence 18, Application US/10484218  
 ; Publication No. US20050059633A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
 ; TITLE OF INVENTION: GLUCANS AND GLUCANSUCRASES DERIVED FROM  
 ; TITLE OF INVENTION: LACTIC ACID BACTERIA  
 ; FILE REFERENCE: 2001-1316  
 ; CURRENT APPLICATION NUMBER: US/10/484,218  
 ; CURRENT FILING DATE: 2004-01-20  
 ; PRIOR APPLICATION NUMBER: PCT/NL02/00495  
 ; PRIOR FILING DATE: 2002-07-22  
 ; PRIOR APPLICATION NUMBER: EP 01202752.0  
 ; PRIOR FILING DATE: 2001-07-20  
 ; PRIOR APPLICATION NUMBER: EP 01202841.1  
 ; PRIOR FILING DATE: 2001-07-25  
 ; NUMBER OF SEQ ID NOS: 30  
 ; SOFTWARE: PatentIn Ver. 3.2  
 ; SEQ ID NO 18  
 ; LENGTH: 1497  
 ; TYPE: PRT  
 ; ORGANISM: Lactobacillus sp.  
 ; FEATURE:  
 ; NAME/KEY: MOD RES  
 ; LOCATION: (284)  
 ; OTHER INFORMATION: Variable amino acid  
 ; FEATURE:  
 ; NAME/KEY: MOD RES  
 ; LOCATION: (1345)  
 ; OTHER INFORMATION: Variable amino acid  
 ; FEATURE:  
 ; NAME/KEY: MOD RES  
 ; LOCATION: (1453)  
 ; OTHER INFORMATION: Variable amino acid  
 US-10-484-218-18

Query Match 44.6%; Score 58; DB 5; Length 1497;  
 Best Local Similarity 52.2%; Pred. No. 19;  
 Matches 12; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 PLDKRSGLNPLIHNSLVDRVDD 23  
 Db 575 PVSQSGLEPELTTSLVNRTGDD 597

RESULT 14  
 US-10-369-493-2411

; Sequence 2411, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 2411  
; LENGTH: 500  
; TYPE: PRT  
; ORGANISM: Schizosaccharomyces pombe  
US-10-369-493-2411

Query Match 41.5%; Score 54; DB 4; Length 500;  
Best Local Similarity 60.0%; Pred. No. 21;  
Matches 12; Conservative 1; Mismatches 3; Indels 4; Gaps 1;

Qy 2 LDKRSGLNPLIHNSLVDREV 21  
||| ||||| |||  
Db 294 LDK---NPLIHNVLTDRSI 309

## RESULT 15

US-09-995-749A-10  
; Sequence 10, Application US/09995749A  
; Patent No. US20020155568A1  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
; FILE REFERENCE: B043388-CIP  
; CURRENT APPLICATION NUMBER: US/09/995,749A  
; CURRENT FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: 09/604,957  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: EPO 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 545  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-995-749A-10

Query Match 40.8%; Score 53; DB 3; Length 545;  
Best Local Similarity 40.6%; Pred. No. 33;  
Matches 13; Conservative 5; Mismatches 6; Indels 8; Gaps 1;

Qy 1 PLDK-----RSLNPLIHNSLVDREVDR 24  
||| ||||| ||| : :  
Db 118 PLEKASNKNEIRSGLEPVTNLSNRSAGK 149

Search completed: February 10, 2006, 23:25:11  
Job time : 15.6673 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:15:51 ; Search time 1.08367 Seconds  
(without alignments)  
302.734 Million cell updates/sec

Title: US-10-797-821-26

Perfect score: 130

Sequence: 1 PLKRSGLNPLIHNLSLVDRDRE 25

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications\_AA\_New.\*

- 1: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB pep.\*
- 2: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB pep.\*
- 3: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB pep.\*
- 4: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB pep.\*
- 5: /cgn2\_6/ptodata/1/pubppaa/US05\_NEW\_PUB pep.\*
- 6: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB pep.\*
- 7: /cgn2\_6/ptodata/1/pubppaa/US11\_NEW\_PUB pep.\*
- 8: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	47	36.2	457	7	US-11-098-686-10234
2	45	34.6	1020	6	US-10-513-786-4
3	45	34.6	1881	7	US-11-040-472-8
4	45	34.6	1900	6	US-10-513-786-3
5	45	34.6	3704	6	US-10-513-786-1
6	44	33.8	323	7	US-11-140-625-9
7	44	33.8	1713	6	US-10-766-317-2
8	44	33.8	1724	6	US-10-766-317-6
9	44	33.8	3333	6	US-10-766-317-4
10	43	33.1	144	6	US-10-793-626-704
11	42	32.3	344	6	US-10-467-657-2616
12	42	32.3	620	6	US-10-793-626-606
13	42	32.3	621	6	US-10-793-626-2614
14	42	32.3	4913	6	US-10-453-372-1142
15	42	32.3	4961	6	US-10-453-372-1132
16	41	31.5	1184	6	US-10-131-826A-412
17	40.5	31.2	1299	6	US-10-821-234-1145
18	40	30.8	452	6	US-10-467-962B-14
19	40	30.8	466	6	US-10-524-647-114
20	40	30.8	729	7	US-11-175-690-270
21	40	30.8	729	7	US-11-175-690-272
22	39.5	30.4	154	6	US-10-793-626-602
23	39	30.0	427	7	US-11-213-368-7
24	38.5	29.6	373	6	US-10-453-372-976
25	38.5	29.6	400	6	US-10-453-372-980

ALIGNMENTS

RESULT 1

US-11-098-686-10234  
; Sequence 10234, Application US/11098686  
; Publication No. US20060024696A1  
; GENERAL INFORMATION:  
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.  
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES  
; FILE REFERENCE: 09531-128001  
; CURRENT APPLICATION NUMBER: US/11/098,686  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: PCT/US03/31318  
; PRIOR FILING DATE: 2003-10-01  
; PRIOR APPLICATION NUMBER: US 60/416,395  
; PRIOR FILING DATE: 2002-10-04  
; NUMBER OF SEQ ID NOS: 11433  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10234  
; LENGTH: 457  
; TYPE: PRT  
; ORGANISM: Lawsonia intracellularis  
US-11-098-686-10234

Query Match 36.2%; Score 47; DB 7; Length 457;  
Best Local Similarity 57.1%; Pred. No. 7.2;  
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 7 GLNPLIHNLSLVDR 20

Db 372 GVTPLVHDPVLVDRE 385

RESULT 2

US-10-513-786-4  
; Sequence 4, Application US/10513786  
; Publication No. US20050260589A1  
; GENERAL INFORMATION:  
; APPLICANT: Bio Control Institute Limited  
; APPLICANT: AGATA, Michio  
; TITLE OF INVENTION: A cereulide synthetase, a gene thereof, and a detection method for  
; FILE REFERENCE: P0202401  
; CURRENT APPLICATION NUMBER: US/10/513,786  
; CURRENT FILING DATE: 2004-11-16  
; PRIOR APPLICATION NUMBER: JP P2002-142398  
; PRIOR FILING DATE: 2002-05-17  
; NUMBER OF SEQ ID NOS: 21



```

Qy 1 PLDKRSGNLNPLIHN--SLVDREVDD 23
Db 34 PIDHSGTGPAVEQAQVSLVERAVKE 58

RESULT 7
US-10-766-317-2
; Sequence 2, Application US/10766317
; Publication No. US20060009630A1
; GENERAL INFORMATION:
; APPLICANT: Marinkovich, M. Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
; FILE REFERENCE: 33828/US/RFT/RMS
; CURRENT APPLICATION NUMBER: US/10/766,317
; CURRENT FILING DATE: 2004-01-27
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 1713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-766-317-2

Query Match 33.8%; Score 44; DB 6; Length 1713;
Best Local Similarity 40.9%; Pred. No. 1e+02;
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 2 LDKRSGNLNPLIHNLSVDREVDD 23
Db 1069 LQKRMWINVDVQNTIIDGEVDF 1090

RESULT 8
US-10-766-317-6
; Sequence 6, Application US/10766317
; Publication No. US20060009630A1
; GENERAL INFORMATION:
; APPLICANT: Marinkovich, M. Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
; FILE REFERENCE: 33828/US/RFT/RMS
; CURRENT APPLICATION NUMBER: US/10/766,317
; CURRENT FILING DATE: 2004-01-27
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6
; LENGTH: 1724
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-766-317-6

Query Match 33.8%; Score 44; DB 6; Length 1724;
Best Local Similarity 40.9%; Pred. No. 1e+02;
Matches 9; Conservative 5; Mismatches 8; Indels 0; Gaps 0;

Qy 2 LDKRSGNLNPLIHNLSVDREVDD 23
Db 1080 LQKRMWINVDVQNTIIDGEVDF 1101

RESULT 9
US-10-766-317-4
; Sequence 4, Application US/10766317
; Publication No. US20060009630A1
; GENERAL INFORMATION:
; APPLICANT: Marinkovich, M. Peter
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR INHIBITING SQUAMOUS CELL CARCINOMA
; FILE REFERENCE: 33828/US/RFT/RMS
; CURRENT APPLICATION NUMBER: US/10/766,317
; CURRENT FILING DATE: 2004-01-27
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4

```

```
QY      5 RSGLNPLIHNSLVDEVD--DRE 25
      : : : : : : : : : : : : : :
Db     127 KQGILFLVHGVEVTPDEIDFRE 149

RESULT 12
US-10-793-626-606
; Sequence 606, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 606
; LENGTH: 620
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-606

Query Match      32.3%; Score 42; DB 6; Length 620;
Best Local Similarity 57.1%; Pred. No. 62;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      8 LNPLIHNSLVDEVD 21
      : : : : : : : : : : : : : :
Db     159 LNPLVDNQYVTKEV 172

RESULT 13
US-10-793-626-2614
; Sequence 2614, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2614
; LENGTH: 621
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-2614

Query Match      32.3%; Score 42; DB 6; Length 621;
Best Local Similarity 57.1%; Pred. No. 62;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      8 LNPLIHNSLVDEVD 21
      : : : : : : : : : : : : : :
Db     159 LNPLVDNQYVTKEV 172

RESULT 14
US-10-453-372-1142
; Sequence 1142, Application US/10453372
```

```
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 1142
; LENGTH: 4913
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-1142

Query Match      32.3%; Score 42; DB 6; Length 4913;
Best Local Similarity 50.0%; Pred. No. 7.4e+02;
Matches 12; Conservative 2; Mismatches 8; Indels 2; Gaps 1;

QY      3 DKRSGLN--PLIHNSLVDEVD 24
      : : : : : : : : : : : : : :
Db     974 DKDSGANDQLYIKSELDRELQDR 997

RESULT 15
US-10-453-372-1132
; Sequence 1132, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHODS
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
```



```
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 1132
; LENGTH: 4961
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-1132

Query Match      32.3%; Score 42; DB 6; Length 4961;
Best Local Similarity 50.0%; Pred. No. 7.5e+02;
Matches 12; Conservative 2; Mismatches 8; Indels 2; Gaps 1;

Qy      3 DKRSGLN--PLIHNSLVDREVDDR 24
Db      1022 DKDSGANDGQLYIKSELDRELQDR 1045

Search completed: February 10, 2006, 23:26:01
Job time : 2.08367 secs
```

**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query		Length	DB	ID	Description
		Match					
1	106	100.0	20	2	AAW34156	Aaw34156	GTF antiG
2	106	100.0	20	2	AAV43336	Aav43336	GTF antiG
3	106	100.0	20	7	AD93647	Ad93647	Streptoco
4	106	100.0	20	9	ADx37270	Adx37270	Streptoco
5	106	100.0	22	2	AAW34161	Aaw34161	GTF antiG
6	103	97.2	498	9	ADW43436	Adw43436	S. sobrin
7	103	97.2	1590	7	AD93657	Ad93657	Streptoco
8	103	97.2	1590	9	ADx37280	Adx37280	Streptoco
9	103	97.2	1592	2	AAK32925	Aak32925	Glucosylt
10	92	86.8	20	2	AAK43697	Aar43697	Gr subseel
11	71	67.0	1499	7	ADC54806	Adc54806	Protein S
12	65	61.3	1375	5	AAU98028	S. mutans	
13	65	61.3	1375	5	AAU79288	Adu79288	Streptoco
14	65	61.3	1375	7	AD93655	Ad93655	Streptoco
15	65	61.3	1375	9	ADx37278	Adx37278	Streptoco
16	62	58.5	380	5	AAU79286	Aau79286	Streptoco
17	62	58.5	726	6	ABU44109	Abu44109	Protein e
18	62	58.5	726	9	AEB91507	Aeb91507	Microbial
19	62	58.5	1475	5	AAU98031	S. mutans	
20	62	58.5	1475	5	AAU98036	Aau98036	S. mutans
21	62	58.5	1475	5	AAU98037	Aau98037	S. mutans
22	62	58.5	1475	5	AAU98040	Aau98040	S. mutans
23	62	58.5	1475	5	AAU98035	S. mutans	
24	62	58.5	1475	5	AAU98033	S. mutans	

CC glucosyltransferase in a mammal. The immune response results in reduction  
 CC of the colonisation or accumulation of mutans streptococcal strains in  
 CC the mammal. Compositions containing AAW34156 specifically interfere with  
 CC the glucan-binding activity of the streptococcal glucosyltransferase. The  
 CC peptides can also be used in vaccines for preventing dental caries in  
 CC mammals. (Updated on 25-MAR-2003 to correct PF field.)

XX Sequence 20 AA;  
 Query Match 100.0%; Score 106; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-10;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGKLRYYDANSQDOAFNKS 20  
 Db 1 DGKLRYYDANSQDOAFNKS 20

RESULT 2  
 AAY43336  
 ID AAY43336 standard; peptide; 20 AA.

XX AC AAY43336;  
 XX 26-JAN-2000 (first entry)  
 DT GTF antigenic epitope.

XX GTF; antigenic epitope; glucosyltransferase; immunogenic epitope; teeth;  
 KW immunogenic composition; streptococcal glucan; antibody response;  
 KW cariogenic bacteria; mutans streptococci; colonisation; caries;  
 KW dental caries; immunisation; therapy.  
 XX Synthetic.

XX WO9952548-A2.  
 XX 21-OCT-1999.  
 XX 09-APR-1999; 99WO-US007828.  
 XX 10-APR-1998; 98US-0081315P.

XX (LEES/) LEES A.  
 PA (TAUB/) TAUBMAN M A.  
 PA (SMIT/) SMITH D J.

XX Lees A, Taubman MA, Smith DJ;  
 XX WPI; 1999-620289/53.

XX Immunogenic compositions for control of dental caries, based on  
 PT Streptococcus mutans components, particularly for vaccination of infants.  
 XX Claim 11; Page 44; 54pp; English.

XX This sequence represents a glucosyltransferase (GTF) immunogenic epitope,  
 CC and can be used in the immunogenic composition of the invention. The  
 CC immunogenic composition (A) comprises at least one streptococcal glucan  
 CC (SG) and at least one component (I) covalently coupled, (in)directly, to  
 CC SG. (A) elicits antibodies (Ab) to both SG and (I). (A), and related  
 CC compositions, are used to induce an antibody response to cariogenic  
 CC bacteria, collectively called "mutans streptococci" (i.e. any of  
 CC Streptococcus mutans, S. rattus, S. cricetus or S. sobrinus), especially  
 CC to prevent colonisation, to reduce numbers of bacteria or to reduce  
 CC incidence of caries, in humans or other animals. The composition can  
 CC therefore be used for the treatment or prevention of caries (particularly  
 CC dental caries). Abs raised against (A) can be used: (a) to detect SG, and  
 CC (b) for passive immunisation, e.g. applied directly to the surface of  
 CC teeth. SG is a poorly immunogenic type 2 T-cell independent antigen but  
 CC when coupled to (I) it produces a conjugate which induces significant  
 CC immunity to both components, resulting in immunological memory and long-  
 CC lasting antibody production against SG, in adults and children

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 106; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-10;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGKLRYYDANSQDOAFNKS 20  
 Db 1 DGKLRYYDANSQDOAFNKS 20

RESULT 3  
 ADD93647  
 ID ADD93647 standard; peptide; 20 AA.

XX AC ADD93647;  
 XX 29-JAN-2004 (first entry)

XX Streptococcus mutans glucosyltransferase glucan-binding domain peptide.  
 DE Streptococcus mutans glucosyltransferase glucan-binding domain peptide.  
 XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.  
 XX Streptococcus mutans.

XX WO2003075845-A2.  
 XX 18-SEP-2003.  
 XX 07-MAR-2003; 2003WO-US006962.  
 XX 07-MAR-2002; 2002US-0363209P.  
 XX 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.  
 XX Smith DJ, Taubman MA;  
 XX WPI; 2003-845091/78.

XX Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.

XX Claim 20; Page 17; 49pp; English.

XX The present sequence is that of a glucan binding domain peptide fragment  
 CC of Streptococcus mutans glucosyltransferase. The peptide can be used in  
 CC immunogenic compositions and subunit vaccines for dental caries. These  
 CC compositions comprise a major histocompatibility complex (MHC) class II  
 CC protein-binding peptide from Streptococcus mutans glucan binding protein  
 CC -B (GbpB) covalently linked with a peptide subunit of a streptococcal  
 CC glucosyltransferase. The compositions are used in a claimed method of  
 CC eliciting production of an antibody in a mammal. Diepitopic or  
 CC multiepitopic polypeptides can be prepared synthetically or by  
 CC recombinant DNA technology. Antibodies raised against MHC class II  
 CC binding fragments of GbpB can be used in passive immunisation.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 106; DB 7; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5.1e-10;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGKLRYYDANSQDOAFNKS 20  
 Db 1 DGKLRYYDANSQDOAFNKS 20

RESULT 4  
 ID ADX37270 standard; peptide; 20 AA.





```

PR 07-MAR-2003; 2003US-00383930.
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX Smith DJ, Taubman MA;
XX WPI; 2005-151644/16.
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX Claim 7; SEQ ID NO 37; 73pp; English.
XX
XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus sobrinus GbpB protein of the invention.
XX
XX Sequence 1590 AA;
SQ
Query Match 97.2%; Score 103; DB 9; Length 1590;
Best Local Similarity 95.0%; Pred. No. 2.1e-07;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DGKLYRYDANSQGQAFNKS 20
DB 1438 DGKRYDYDANSQGQAFNKS 1457
RESULT 9
AAR32925
ID AAR32925 standard; protein; 1592 AA.
XX
XX AAR32925;
XX
XX 28-JUN-1993 (first entry)
XX
XX Glucosyltransferase I.
DE
XX GT-1; Streptococcus; dental; caries.
XX
XX Streptococcus sobrinus.
XX
XX JP05023188-A.
XX
XX 02-FEB-1993.
XX
XX 25-JUL-1991; 91JP-00186592.
XX
XX 25-JUL-1991; 91JP-00186592.
XX
XX (KATO/) KATO K.
PA (FUKU/) FUKUI I.
XX
XX WPI; 1993-079449/10.
DR N-PSDB; AAQ37760.
XX
XX DNA sequence Glucosyltransferase-I - comprises Streptococcus sobrinus
PT DNA sequence with at least one nucleotide added or deleted.
XX
XX Claim 13; Page 15; 29pp; Japanese.
XX
XX The DNA sequence from Streptococcus sobrinus strain 6715 encodes
CC Glucosyltransferase-I (and mutants). The DNA was obtd. by treating S.
CC sobrinus 6715 with mutanolysin, extracting the chromosomal DNA, partially
CC digesting with Sau3AI and fractionating on agarose gel. The 3-5 kbp
CC fragment was ligated into pUC18 and E. coli JM109 transformed with it. A
CC GT-1 expressing clone was isolated and sequenced. The clone may be used

```

---

```

CC in the development of a drug for dental caries
XX
XX Sequence 1592 AA;
Query Match 97.2%; Score 103; DB 2; Length 1592;
Best Local Similarity 95.0%; Pred. No. 2.1e-07;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 DGKLYRYDANSQGQAFNKS 20
DB 1440 DGKRYDYDANSQGQAFNKS 1459
RESULT 10
AAR43697
ID AAR43697 standard; peptide; 20 AA.
XX
XX AAR43697;
XX
XX 25-MAR-2003 (revised)
DT 20-MAY-1994 (first entry)
XX
XX GT subsequence.
DE
XX GT; glucosyltransferase; vaccine; T-cell; B-cell; reaction;
KW immunoresponse; peptidyl core matrix; dental caries; diphtheria; tetanus;
KW measles; polio.
XX
XX Synthetic.
XX
XX WO9322341-A1.
XX
XX 11-NOV-1993.
XX
XX 30-APR-1993; 93WO-US004094.
XX
XX 01-MAY-1992; 92US-00877295.
XX
XX (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
XX
XX Taubman MA, Smith DJ;
XX WPI; 1993-368721/46.
XX
XX Synthetic glucosyl-transferase peptide(s) - provoke T- and B-cell
PT reactions in mammals, and are effective against dental caries.
XX
XX Claim 3; Page 24; 38pp; English.
XX
XX The sequences (AAR43694-98) are subsequences from GT, they provoke T- and
CC B-cell responses in mammals. The can be used to produce a vaccine
CC comprising 2 of these sequences attached to a peptidyl core matrix. They
CC are useful in producing T-cell responses to the GT enzyme causing a
CC reduction of colonisation or accumulation of mutans streptococcal strains
CC in mammals. The vaccines can be used in preventing dental caries.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 20 AA;
SQ
Query Match 86.8%; Score 92; DB 2; Length 20;
Best Local Similarity 90.0%; Pred. No. 9.9e-08;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 DGKLYRYDANSQGQAFNKS 20
DB 1 DGKLYRYDANSQGQAFNKS 20
RESULT 11
ADC54806
ID ADC54806 standard; protein; 1499 AA.
XX
XX ADC54806;
AC

```

XX 18-DEC-2003 (first entry)  
 XX Protein Seq ID11 related to L mesenteroides dextran sucrose protein.  
 DE dextran sucrose; active centre zone; glucan; polysaccharide; dextran;  
 XX D-glucose; starch; cellulose; glucan manufacture; transduction;  
 KW enzyme-reaction product.  
 KW Unidentified.  
 XX OS  
 XX JP2003111590-A.  
 XX 15-APR-2003.  
 XX 03-OCT-2001; 2001JP-00307067.  
 XX 03-OCT-2001; 2001JP-00307067.  
 XX (DOKU-) DOKURITSU GYOSEI HOJIN SHOKUJIN SOGO KEN.  
 XX WPI; 2003-735670/70.  
 DR P-PSDB; ADC54814.  
 XX Novel modified dextran sucrose which exchanges one site of active center  
 PT zone of dextran sucrose for active center zone of different types of  
 PT dextran sucrose, useful for manufacturing glucan.  
 XX Example 1; SEQ ID NO 11; 28pp; Japanese.  
 XX This invention relates to a modified dextran sucrose (DS) exchanging one  
 CC site of the active centre zone of a dextran sucrose for the active centre  
 CC zone of different types of dextran sucrose. Glucan is a polysaccharide  
 CC (for example dextran) which uses D-glucoses, such as a starch and a  
 CC cellulose, as a structural unit. The modified enzyme of the invention is  
 CC useful in the manufacture of glucan. The selection of the active centre  
 CC area of DS which carries out transduction, enables changes in structure  
 CC and character of an enzyme-reaction product and their application to  
 CC various uses. The present sequence is that of a protein which is related  
 CC to the leuconostoc mesenteroides dextran sucrose protein and which was  
 CC used during the exemplification of the invention.  
 XX Sequence 1499 AA;  
 SQ Query Match 67.0%; Score 71; DB 7; Length 1499;  
 Best Local Similarity 66.7%; Pred. No. 0.033;  
 Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 DGKLYRYDANSQDAFNK 18  
 |||:|||||:|:  
 DB 1418 DGKRYDYDADSGDMVTNR 1435  
 RESULT 12  
 AAU98028  
 ID AAU98028 standard; protein; 1375 AA.  
 XX AAU98028;  
 AC AAU98028;  
 XX 27-AUG-2002 (first entry)  
 XX S. mutans glucosyltransferase GTFC.  
 DE Glucosyltransferase; GTFC; transgenic plant; paper sizing;  
 XX coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture.  
 XX Streptococcus mutans.  
 OS  
 XX US2002031826-A1.  
 XX 14-MAR-2002.  
 PD AAU79288  
 XX ID AAU79288 standard; protein; 1375 AA.

PF 19-DEC-2000; 2000US-00740274.  
 XX 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 PA Nichols SE;  
 XX WPI; 2002-414332/44.  
 DR N-PSDB; ABK52939.  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Disclosure; Page 30-33; 44pp; English.  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I48V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents GTFC  
 XX  
 SQ Sequence 1375 AA;  
 Query Match 61.3%; Score 65; DB 5; Length 1375;  
 Best Local Similarity 60.0%; Pred. No. 0.29;  
 Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 DGKLYRYDANSQDAFNKSV 20  
 |||||:|:|:|:  
 DB 1227 DGKLYRFDRDSGNQISNRFV 1246  
 RESULT 13  
 AAU79288  
 ID AAU79288 standard; protein; 1375 AA.



XX AAU79288;  
 AC 13-AUG-2002 (first entry)  
 DT Streptococcus mutans monoclonal antibody-related protein #5.  
 DE  
 XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;  
 KW Glucosyl transferase-B; immunotherapy.  
 XX Streptococcus mutans.  
 OS  
 XX JP2002114709-A.  
 PN  
 XX 16-APR-2002.  
 PD  
 XX 04-OCT-2000; 2000JP-00304889.  
 PF  
 XX 04-OCT-2000; 2000JP-00304889.  
 PR  
 XX (UYN1-) UNIV NIPPON.  
 PA  
 XX WPI; 2002-448885/48.  
 DR  
 XX Anti-carries agent composed of a monoclonal antibody against an inhibitory  
 PT enzyme against water insoluble glucan synthetase of glucosyl transferase-  
 FT B (GTF-B) of Streptococcus mutans.  
 PR  
 XX Disclosure; Page 22-25; 28pp; Japanese.  
 PS  
 XX The invention relates to a monoclonal antibody against dental caries and  
 CC an anti-carries agent composed of a monoclonal antibody produced by  
 CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)  
 CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having  
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
 CC transferase-B. The monoclonal antibody specifically inhibits water  
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl  
 CC transferase-B and is used in the immunotherapy of dental caries. This  
 CC sequence represents a Streptococcus mutans monoclonal antibody-related  
 CC protein.  
 CC  
 XX SQ Sequence 1375 AA;  
 Query Match 61.3%; Score 65; DB 5; Length 1375;  
 Best Local Similarity 60.0%; Pred. No. 0.29;  
 Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 DGKLYRYDANSQDAFNKSV 20  
 DB 1227 DGKLYRFDSDGNQISNRFV 1246  
 RESULT 14  
 ADD93655  
 ID ADD93655 standard; protein; 1375 AA.  
 XX  
 AC ADD93655;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX Streptococcus mutans glucosyltransferase-C.  
 DE  
 XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.  
 KW  
 XX Streptococcus mutans.  
 OS  
 XX WO2003075845-A2.  
 PN  
 XX 18-SEP-2003.  
 PD  
 XX 07-MAR-2003; 2003WO-US006962.  
 PF  
 XX 07-MAR-2002; 2002US-0363209P.  
 PR

PR 08-AUG-2002; 2002US-0402483P.  
 XX (FORSY-) FORSYTH INST.  
 PA  
 XX Smith DJ, Taubman MA;  
 PI  
 XX WPI; 2003-845091/78.  
 DR  
 XX Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 FT class II protein.  
 PR  
 XX Claim 16; Page 13; 49pp; English.  
 PS  
 XX The present sequence is the protein sequence of Streptococcus mutans  
 CC glucosyltransferase-C (GTF-C). Peptide fragments of GTF-C, especially  
 CC from the catalytic domain of the polypeptide, can be used in immunogenic  
 CC compositions and subunit vaccines for dental caries. These compositions  
 CC comprise a major histocompatibility complex (MHC) class II protein-  
 CC binding peptide from S. mutans glucan binding protein-B (GbpB)  
 CC covalently linked with a peptide fragment of a streptococcal  
 CC glucosyltransferase. The compositions are used in a claimed method of  
 CC eliciting production of an antibody in a mammal. Diepitopic or  
 CC multi-epitopic polypeptides can be prepared synthetically or by  
 CC recombinant DNA technology. Antibodies raised against MHC class II  
 CC binding fragments of GbpB can be used in passive immunisation.  
 XX  
 SQ Sequence 1375 AA;  
 Query Match 61.3%; Score 65; DB 7; Length 1375;  
 Best Local Similarity 60.0%; Pred. No. 0.29;  
 Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 DGKLYRYDANSQDAFNKSV 20  
 DB 1227 DGKLYRFDSDGNQISNRFV 1246  
 RESULT 15  
 ADX37278  
 ID ADX37278 standard; protein; 1375 AA.  
 XX  
 AC ADX37278;  
 XX  
 DT 21-APR-2005 (first entry)  
 XX Streptococcus mutant glucan binding protein B variant #7.  
 DE  
 XX immunogenicity; immune stimulation; glucan binding protein-B;  
 KW microparticle; major histocompatibility complex; tooth disease.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PN US2005031633-A1.  
 XX  
 PD 10-FEB-2005.  
 PF  
 XX 09-MAR-2004; 2004US-00797821.  
 XX  
 DT 13-APR-1998; 98US-0081550P.  
 XX  
 PR 08-JAN-1999; 99US-0115142P.  
 PR 12-APR-1999; 99US-00290049.  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 PR 07-MAR-2003; 2003US-00383930.  
 XX  
 XX (SMIT/) SMITH D J.  
 PA (TAUB/) TAUBMAN M A.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX WPI; 2005-151644/16.  
 DR  
 XX

PT New composition comprising a fragment of a glucan binding protein-B  
 PT (GbpB) that binds to MHC class II protein, and a biocompatible  
 PT microparticle, useful for producing an antibody (claimed) for immunizing  
 PT mammals against dental caries.

XX Claim 7; SEQ ID NO 35; 73pp; English.

XX The invention relates to a composition comprising a fragment of a glucan  
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II  
 CC protein. The composition is useful for producing an antibody for  
 CC immunizing mammals against dental caries. This sequence corresponds to a  
 CC Streptococcus mutans GbpB protein of the invention.

XX Sequence 1375 AA;

Query Match 61.3%; Score 65; DB 9; Length 1375;  
 Best Local Similarity 60.0%; Pred. No. 0.29;  
 Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 DGKLYVDANSQDAFNKSV 20  
 Db 1227 DGKLYFDSDGNQISNRFV 1246

Search completed: February 10, 2006, 22:19:43  
 Job time : 14.4758 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:20:08 ; Search time 2.35887 Seconds  
(without alignments)  
815.787 Million cell updates/sec

Title: US-10-797-821-27

Perfect score: 106

Sequence: 1 DGKLRYYDANSGDQAFNKS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 80.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	103	97.2	1592	2 A38175	glucosyltransferas
2	65	61.3	1375	2 JT0345	dextranucrase (EC
3	64	60.4	1449	2 T30857	glucosyltransferas
4	64	60.4	1449	2 T30552	glucosyltransferas
5	62	58.5	1475	2 B33135	gtfB protein precu
6	57	53.8	1599	2 S22737	glucosyltransferas
7	56	52.8	1431	2 A45866	dextranucrase (EC
8	52	49.1	1122	2 T18346	MGC1 protein precu
9	50	47.2	1290	2 JC5473	dextranucrase (EC
10	50	47.2	1577	2 T30858	glucosyltransferas
11	48	45.3	1365	2 A41483	Glucosyltransferas
12	47	44.3	695	2 S51433	MDL1 protein - yea
13	46	43.4	26	2 S64715	formin binding pro
14	46	43.4	1098	2 T08599	probable transcrip
15	46	43.4	1508	2 T31098	probable dextranu
16	45	42.5	558	2 T18467	hypothetical prote
17	45	42.5	595	2 T39228	beta-transducin -
18	45	42.5	1230	2 F82707	pilY1 protein homo
19	45	42.5	1217	2 P82857	pilY1 protein homo
20	44.5	42.0	378	2 G84093	integrase (phage-r
21	44.5	42.0	531	2 AH3378	cell wall degradat
22	44	41.5	325	2 T09613	probable GTP-bindi
23	44	41.5	859	2 S69700	hypothetical prote
24	44	41.5	1020	2 T18260	1-phosphatidylinos
25	43	40.6	329	2 A55221	dextranase inhibito
26	43	40.6	336	1 C41659	benzoate 1,2-dioxy
27	43	40.6	380	2 S51797	vasodilator-stimul
28	43	40.6	383	2 T26902	hypothetical prote
29	43	40.6	406	2 T24301	hypothetical prote

30	43	40.6	464	2 T34332	hypothetical prote
31	43	40.6	507	2 D64575	hypothetical prote
32	43	40.6	1033	2 S54506	probable membrane
33	43	40.6	1965	2 T33216	hypothetical prote
34	42.5	40.1	370	2 B83128	probable hydrolase
35	42.5	40.1	372	2 G96933	HD-GYP hydrolase d
36	42.5	40.1	1132	2 AF2481	hypothetical prote
37	42	39.6	173	2 T24207	hypothetical prote
38	42	39.6	210	2 T27471	hypothetical prote
39	42	39.6	243	2 H69224	hypothetical prote
40	42	39.6	286	2 D84591	hypothetical prote
41	42	39.6	301	2 D90224	spermidine synthas
42	42	39.6	377	2 B69828	hypothetical prote
43	42	39.6	578	2 T15958	hypothetical prote
44	42	39.6	645	2 T05991	hypothetical prote
45	42	39.6	772	2 JC4696	alpha.alpha-trehal

ALIGNMENTS

RESULT 1

A38175

glucosyltransferase precursor - Streptococcus sobrinus

C;Species: Streptococcus sobrinus

C;Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Mar-2004

C;Accession: A38175

R;Abo, H.; Mtsumura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.

J. Bacteriol. 173, 989-996, 1991

A;Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus

A;Reference number: A38175; MUID:91123227; PMID:1704006

A;Accession: A38175

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1592 <ABO>

A;Cross-references: UNIPARC:UPI000012ECB2; GB:D90213; NID:g217032; PIDN:BAA14241.1; PID:

F;1093-1112/Domain: cpl repeat homology <CP1>

F;1222-1241/Domain: cpl repeat homology <CP2>

F;1287-1306/Domain: cpl repeat homology <CP3>

F;1330-1351/Domain: cpl repeat homology <CP4>

F;1352-1371/Domain: cpl repeat homology <CP5>

F;1402-1420/Domain: cpl repeat homology <CP6>

F;1465-1484/Domain: cpl repeat homology <CP7>

F;1513-1532/Domain: cpl repeat homology <CP8>

Query Match 97.2% Score 103; DB 2; Length 1592;

Best Local Similarity 95.0% Pred. No. 5.le-08;

Matches 13; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGKLRYYDANSGDQAFNKS 20

Db 1440 DGKRVYYDANSGDQAFNKS 1459

RESULT 2

JT0345

dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)

N;Alternate names: sucrose 6-glucosyltransferase

C;Species: Streptococcus mutans

C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004

C;Accession: JT0345; C33135

R;Ueda, S.; Shiroza, T.; Kuramitsu, H.K.

Gene 69, 101-109, 1988

A;Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.

A;Reference number: JT0345; MUID:89137980; PMID:2976010

A;Accession: JT0345

A;Molecule type: DNA

A;Residues: 1-1375 <UED>

A;Cross-references: UNIPROT:P13470; UNIPARC:UPI0000155515

A;Experimental source: GS-5

R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987

A;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.

A;Reference number: A33135; MUID:87308013; PMID:3040685  
A;Accession: C33135  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-349 <SHI>  
A;Cross-references: UNIPARC:UPI000014E25D; GB:M17361  
C;Genetics:  
A;Gene: gtfC  
C;Function:  
A;Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
A;Keywords: duplication; glycosyltransferase; hexosyltransferase  
F;1-34/Domain: signal sequence #status predicted <SIG>  
F;35-1375/Product: glycosyltransferase #status predicted <MAT>  
F;1126-1145/Domain: cpl repeat homology <CP1>  
F;1253-1272/Domain: cpl repeat homology <CP2>  
F;1318-1337/Domain: cpl repeat homology <CP3>  
  
Query Match 61.3%; Score 65; DB 2; Length 1375;  
Best Local Similarity 60.0%; Pred. No. 0.053;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 1 DGKLRYYDANSQDAFNKSV 20  
Db 1227 DGKLRYYDRDSGNQISNRFV 1246  
|||||:|||||:|:  
|:|:|:|:|:|:|:|:  
  
RESULT 3  
T30857  
glucosyltransferase - Streptococcus salivarius  
C;Species: Streptococcus salivarius  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
A;Accession: T30857  
A;Reference number: 220909; MUID:95122197; PMID:7822030  
A;Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri  
A;Reference number: 220909; MUID:95122197; PMID:7822030  
A;Accession: T30857  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1449 <SIM>  
A;Cross-references: UNIPROT:Q55264; UNIPARC:UPI00000B166E; EMBL:L35495; NID:G662378; PID  
C;Genetics:  
A;Gene: gtfL  
  
Query Match 60.4%; Score 64; DB 2; Length 1449;  
Best Local Similarity 61.1%; Pred. No. 0.081;  
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 DGKLRYYDANSQDAFNK 18  
Db 1285 NGAIRYYDANSGEWARR 1302  
:|:|||||:|:  
|:|:|:|:|:|:|:|:  
  
RESULT 4  
T30552  
glucosyltransferase N - Streptococcus salivarius (fragment)  
C;Species: Streptococcus salivarius  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T30552  
R;Jaffe, R.I.  
submitted to the EMBL Data Library, February 1998  
A;Description: Streptococcus salivarius V1477 gtfN.  
A;Reference number: 220854  
A;Accession: T30552  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1449 <JAF>  
A;Cross-references: UNIPROT:O68542; UNIPARC:UPI00000B10FD; EMBL:AF049609; NID:g2935545;  
C;Genetics:  
A;Gene: gtfN  
  
Query Match 60.4%; Score 64; DB 2; Length 1449;  
Best Local Similarity 61.1%; Pred. No. 0.081;

Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 1 DGKLRYYDANSQDAFNK 18  
Db 1285 NGAIRYYDANSGEWARR 1302  
:|:|||||:|:  
|:|:|:|:|:|:|:|:  
  
RESULT 5  
B33135  
gtfB protein precursor - Streptococcus mutans  
C;Species: Streptococcus mutans  
C;Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 09-Jul-2004  
C;Accession: B33135; A33128  
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A;Reference number: A33135; MUID:87308013; PMID:3040685  
A;Accession: B33135  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1475 <SHI>  
A;Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:g153639; PIDN:  
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
submitted to the Protein Sequence Database, September 1990  
A;Reference number: A33128  
A;Accession: A33128  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-171,173-641,'N',643-1475 <SH2>  
A;Cross-references: UNIPARC:UPI000017AC5E  
A;Experimental source: strain GS-5  
F;1096-1115/Domain: cpl repeat homology <CP1>  
F;1224-1243/Domain: cpl repeat homology <CP2>  
F;1289-1308/Domain: cpl repeat homology <CP3>  
F;1354-1373/Domain: cpl repeat homology <CP4>  
F;1419-1438/Domain: cpl repeat homology <CP5>  
  
Query Match 58.5%; Score 62; DB 2; Length 1475;  
Best Local Similarity 57.9%; Pred. No. 0.17;  
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
  
Qy 2 GKLRYYDANSQDAFNKSV 20  
Db 1264 GRISYYDGNSGQIRNRFV 1282  
|:|:|||||:|:  
|:|:|:|:|:|:|:|:  
  
RESULT 6  
S22737  
glucosyltransferase (EC 2.4.1.-) S - Streptococcus salivarius  
C;Species: Streptococcus salivarius  
C;Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 09-Jul-2004  
C;Accession: S22737; S28810; E44811; S22727  
R;Jacques, N.  
submitted to the EMBL Data Library, March 1992  
A;Reference number: S22726  
A;Accession: S22737  
A;Molecule type: DNA  
A;Residues: 1-1599 <JAC>  
A;Cross-references: UNIPROT:Q00599; UNIPARC:UPI000008BEF34; EMBL:Z11872; NID:g47530; PIDN:  
R;Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A;Title: Molecular characterization of a cluster of at least two glucosyltransferase gene  
A;Reference number: A44811; MUID:92148377; PMID:1838391  
A;Accession: S28810  
A;Molecule type: DNA  
A;Residues: 1-51 <GIF>  
A;Cross-references: UNIPARC:UPI000017027C; EMBL:Z11873  
C;Genetics:  
A;Gene: gtfK  
A;Keywords: glycosyltransferase; hexosyltransferase  
F;1456-1475/Domain: cpl repeat homology <CPR>

Query Match 53.8%; Score 57; DB 2; Length 1599;  
Best Local Similarity 45.0%; Pred. No. 1.2;  
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DGKLRYYDANSQDAFNKSV 20  
Db 1560 DGSISYYDVHTGKAINRWV 1579  
|||:||||:||||:|

RESULT 7  
A45866  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans  
C;Species: Streptococcus mutans  
C;Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 15-Mar-2004  
C;Accession: A45866  
R;Honda, O.; Kato, C.; Kuramitsu, H.K.  
J. Gen. Microbiol. 136, 2099-2105, 1990  
A;Title: Nucleotide sequence of the Streptococcus mutans gtfd gene encoding the glucosyltransferase  
A;Reference number: A45866; MUID:91100958; PMID:2148600  
A;Accession: A45866  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1431 <HON>  
A;Cross-references: UNIPARC:UPI000017AC5C; GB:M29296  
C;Keywords: glycosyltransferase; hexosyltransferase  
F;181-201/Domain: cpl repeat homology <CP1>  
F;1127-1146/Domain: cpl repeat homology <CP2>  
F;1192-1211/Domain: cpl repeat homology <CP3>  
F;1257-1276/Domain: cpl repeat homology <CP4>  
F;1277-1297/Domain: cpl repeat homology <CP5>  
F;1321-1340/Domain: cpl repeat homology <CP8>  
F;1341-1361/Domain: cpl repeat homology <CP6>  
F;1385-1404/Domain: cpl repeat homology <CP7>

Query Match 52.8%; Score 56; DB 2; Length 1431;  
Best Local Similarity 55.6%; Pred. No. 1.5;  
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DGKLRYYDANSQDAFNK 18  
Db 1231 DGKKRYFDDSGNVAVR 1248  
|||:||||:||||:|

RESULT 8  
T18346  
MGC1 protein precursor - Mycoplasma gallisepticum  
C;Species: Mycoplasma gallisepticum  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T18346  
R;Keeler Jr., C.L.; Hnatow, L.L.; Whetzel, P.L.; Dohme, J.E.  
Infect. Immun. 64, 1541-1547, 1996  
A;Title: Cloning and characterization of a putative cytoadhesin gene (mgc1) from Mycoplasma  
A;Reference number: Z18881; MUID:96201559; PMID:8613358  
A;Accession: T18346  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1122 <KEE>  
A;Cross-references: UNIPROT:Q49379; UNIPARC:UPI00001255FB; EMBL:U34842; NID:g1022701; PID:g1022701  
C;Genetics:  
A;Gene: mgc1  
A;Genetic code: SGC3

Query Match 49.1%; Score 52; DB 2; Length 1122;  
Best Local Similarity 50.0%; Pred. No. 5.1;  
Matches 12; Conservative 3; Mismatches 3; Indels 6; Gaps 1;

Qy 2 GKLRYYDAN-----SGDQAFNKS 19  
Db 108 KGKGYFDANNKLTAFSGDVSFNVS 131  
|||:||||:||||:|

RESULT 9  
JC5473

dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides  
C;Species: Leuconostoc mesenteroides  
C;Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 09-Jul-2004  
C;Accession: JC5473  
R;Monchois, V.; Willemot, R.M.; Remaud-Simeon, M.; Croux, C.; Monsan, P.  
Gene 182, 23-32, 1996  
A;Title: Cloning and sequencing of a gene coding for a novel dextranucrase from Leuconostoc  
A;Reference number: JC5473; MUID:97136686; PMID:8982063  
A;Accession: JC5473  
A;Status: nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-1290 <MON>  
A;Cross-references: UNIPROT:Q48756; UNIPARC:UPI000017ABDF; GB:U38181  
C;Comment: This enzyme catalyzes the transfer of D-glucopyranosyl units from sucrose onto  
C;Genetics:  
A;Gene: dsra  
C;Keywords: glycosyltransferase; hexosyltransferase  
F;78-870/Domain: catalytic #status predicted <CAT>  
F;922-1290/Domain: glucan-binding #status predicted <GCB>

Query Match 47.2%; Score 50; DB 2; Length 1290;  
Best Local Similarity 45.0%; Pred. No. 12;  
Matches 9; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

Qy 1 DGKLRYYDANSQDAFNKSV 20  
Db 1208 NGIMRYDADTGTGLVRNFV 1227  
|||:||||:||||:|

RESULT 10  
T30858  
glucosyltransferase - Streptococcus salivarius  
C;Species: Streptococcus salivarius  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T30858  
R;Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A;Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for pri  
A;Reference number: Z20909; MUID:95122197; PMID:7822030  
A;Accession: T30858  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1577 <SIM>  
A;Cross-references: UNIPROT:Q5265; UNIPARC:UPI00000B8087; EMBL:L35928; NID:g662380; PID:g662380  
C;Genetics:  
A;Gene: gtfm

Query Match 47.2%; Score 50; DB 2; Length 1577;  
Best Local Similarity 47.4%; Pred. No. 15;  
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 2 GKLRYYDANSQDAFNKSV 20  
Db 1497 GKCYFDAHTGEQVVRVF 1515  
|||:||||:||||:|

RESULT 11  
A41483  
glucosyltransferase (EC 2.4.1.-) gtfs precursor - Streptococcus sobrinus  
C;Species: Streptococcus sobrinus  
C;Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 15-Mar-2004  
C;Accession: A41483  
R;Gillmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
Infect. Immun. 58, 2452-2458, 1990  
A;Title: Analysis of the Streptococcus downei gtfs gene, which specifies a glucosyltransferase  
A;Reference number: A41483; MUID:90316665; PMID:2142479  
A;Accession: A41483  
A;Molecule type: DNA  
A;Residues: 1-1365 <GIL>  
A;Cross-references: UNIPARC:UPI000012BCB6; GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:g153652  
C;Genetics:  
A;Gene: gtfs  
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 45.3%; Score 48; DB 2; Length 1365;  
Best Local Similarity 61.5%; Pred. No. 28;  
Matches 8; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 DGKLRYYDANSQD 13  
                  :|||||:|:  
DB 1328 NGNLRVFDADSGN 1340

## RESULT 12

S51433  
MDL1 protein - yeast (Saccharomyces cerevisiae)  
N;Alternate names: protein L9470.3; protein YLR188w  
C;Species: Saccharomyces cerevisiae  
C;Date: 23-Feb-1995 #sequence\_revision 11-Aug-1995 #text\_change 05-Oct-2004  
C;Accession: S51433; S42681  
R;Wohldmann, P.  
submitted to the EMBL Data Library, November 1994  
A;Description: The sequence of S. cerevisiae cosmid 9470.  
A;Reference number: S51414  
A;Accession: S51433  
A;Molecule type: DNA  
A;Residues: 1-695 <MOH>  
A;Cross-references: UNIPROT:P33310; UNIPARC:UPI000012EE63; EMBL:U17246; NID:G577192; PDB:1YLR188w  
R;Dean, M.; Allikmets, R.; Gerrard, B.; Stewart, C.; Kistler, A.; Shafer, B.; Michaelis, E.; Yeast 10, 377-383, 1994  
A;Title: Mapping and sequencing of two yeast genes belonging to the ATP-binding cassette family  
A;Reference number: S42681  
A;Accession: S42681  
A;Molecule type: DNA  
A;Residues: 1-21, 'G', 23-149, 'L', 151-266, 'WP', 269-312, 'KO', 315-416, 'SSDYVKGSCVISPKTH', 417-481  
A;Cross-references: UNIPARC:UPI0000148B20; EMBL:L16958; NID:G295583; PIDN:AAA20681.1; PDB:1YLR188w  
C;Genetics:  
A;Gene: SGD:MDL1  
A;Cross-references: SGD:S0004178; MIPS:YLR188w  
A;Map position: 12R  
C;Keywords: ATP; nucleotide binding; P-loop; transmembrane protein  
P;103-119/Domain: transmembrane #status predicted <TM1>  
F;161-177/Domain: transmembrane #status predicted <TM2>  
F;256-272/Domain: transmembrane #status predicted <TM3>  
F;450-648/Domain: ATP-binding cassette homology <ABC>  
F;467-474/Region: nucleotide-binding motif A (P-loop)

Query Match 44.3%; Score 47; DB 2; Length 695;  
Best Local Similarity 69.2%; Pred. No. 19;  
Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 LRYVDANSQDAF 16  
                  |||||  
DB 481 LRYVDNSGSI 493

## RESULT 13

S64715  
formin binding protein 23 - mouse (fragment)  
C;Species: Mus musculus (house mouse)  
C;Date: 28-Oct-1996 #sequence\_revision 13-Mar-1997 #text\_change 09-Jul-2004  
C;Accession: S64715  
R;Chan, D.C.; Bedford, M.T.; Leder, P.  
EMBO J. 15, 1045-1054, 1996  
A;Title: Formin binding proteins bear WWP/WW domains that bind proline-rich peptides and regulate the actin cytoskeleton  
A;Reference number: S64711; MUID:96183189; PMID:8605874  
A;Accession: S64715  
A;Status: preliminary; nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-26 <CHA>  
A;Cross-references: UNIPROT:Q61050; UNIPARC:UPI0000069E8; EMBL:U40748; NID:g1255026; PDB:1YLR188w  
C;Superfamily: formin binding protein; WW repeat homology  
F;1-26/Domain: WW repeat homology (fragment) <WW1>

Query Match 43.4%; Score 46; DB 2; Length 26;  
Best Local Similarity 44.4%; Pred. No. 0.86;

Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;  
QY 1 DGKLRYYDANSQDAFNK 18  
                  |||||:|:  
DB 8 DGKYYYNARTRESAWTK 25

## RESULT 14

T08599  
probable transcription factor CA150 - human  
C;Species: Homo sapiens (man)  
C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
C;Accession: T08599  
R;Sune, C.; Hayashi, T.; Liu, Y.; Lane, W.S.; Young, R.A.; Garcia-Blanco, M.A.  
Mol. Cell. Biol. 17, 6029-6039, 1997  
A;Title: CA150, a nuclear protein associated with the RNA polymerase II holoenzyme, is involved in the regulation of gene expression  
A;Reference number: Z16449; MUID:97459702; PMID:9315662  
A;Accession: T08599  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-1098 <SUN>  
A;Cross-references: UNIPROT:O14776; UNIPARC:UPI0000072769; EMBL:AF017789; NID:g2460123; PDB:1YLR188w  
A;Experimental source: cell line HeLa  
C;Genetics:  
A;Gene: CA150  
C;Function:  
A;Description: HIV-1 Tat transcriptional coactivator  
C;Keywords: nucleus; transcription regulation  
F;131-168/Domain: WW repeat homology <WW1>  
F;429-466/Domain: WW repeat homology <WW2>  
F;527-565/Domain: WW repeat homology <WW3>

Query Match 43.4%; Score 46; DB 2; Length 1098;  
Best Local Similarity 44.4%; Pred. No. 46;  
Matches 8; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 DGKLRYYDANSQDAFNK 18  
                  |||||:|:  
DB 144 DGKYYYNARTRESAWTK 161

## RESULT 15

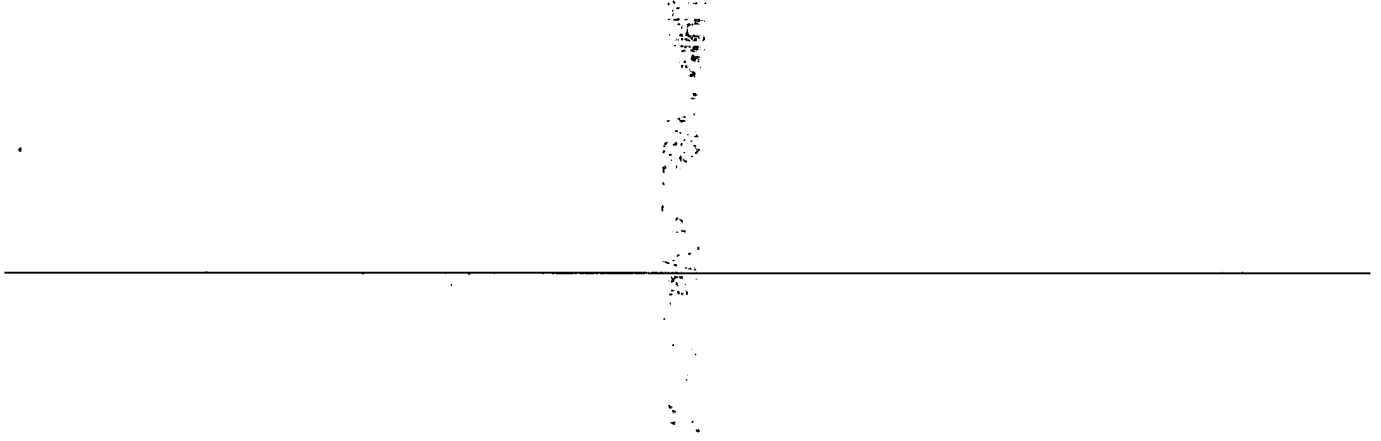
T31098  
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides  
C;Species: Leuconostoc mesenteroides  
C;Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T31098  
R;Monchois, V.; Remaud-Simeon, M.; Monsan, P.; Willemot, R.M.  
FEMS Microbiol. Lett. 159, 307-315, 1998  
A;Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (DSF) from Leuconostoc mesenteroides  
A;Reference number: Z20981; MUID:98164374; PMID:9503626  
A;Accession: T31098  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1508 <MON>  
A;Cross-references: UNIPROT:Q52224; UNIPARC:UPI000008B69B; EMBL:AF030129; NID:g2766611; PDB:1YLR188w  
A;Experimental source: strain NRRL B-1299  
C;Genetics:  
A;Gene: dsrB  
C;Function:  
A;Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
C;Keywords: glycosyltransferase; hexosyltransferase

Query Match 43.4%; Score 46; DB 2; Length 1508;  
Best Local Similarity 50.0%; Pred. No. 64;  
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 DGKLRYYDANSQDQ 14  
                  |||||:|:  
DB 270 DGQVRYFDQESQGE 283

Search completed: February 10, 2006, 22:33:18

Job time : 4.48387 secs



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:07:53 ; Search time 13.7298 Seconds  
(without alignments)  
1027.730 Million cell updates/sec

Title: US-10-797-821-27  
Perfect score: 106  
Sequence: 1 DGLKRYDANSQDAFNKSV 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	1597	1	GTF1_STRDO
2	103	97.2	1590	2	Q55263_9STRE
3	103	97.2	1590	2	Q59983_9STRE
4	103	97.2	1592	1	GTF2_STRDO
5	71	67.0	1454	2	Q69A94_LEUME
6	65	61.3	1451	2	Q8VUH3_STRMU
7	65	61.3	1455	1	GTF3_STRMU
8	65	61.3	1462	1	GTFD_STRMU
9	64	60.4	1449	2	Q68542_STRSL
10	64	60.4	1449	2	Q55264_STRSL
11	64	60.4	1575	2	Q9LCH3_STROR
12	63	59.4	1476	1	GTFB_STRMU
13	63	59.4	1577	2	Q54178_STRGN
14	62	58.5	726	2	Q8DUW9_STRMU
15	57	53.8	421	1	TOLB_WOLSU
16	57	53.8	1561	2	Q5SBM8_9LACO
17	57	53.8	1599	2	Q00599_STRSL
18	54	50.9	1338	2	Q9WXJ4_9STRE
19	53	50.0	243	2	Q6BKR7_DEBHA
20	53	50.0	1512	2	Q9WXJ5_9STRE
21	53	50.0	1595	2	Q5SBM3_LACSC
22	52	49.1	1122	1	ADP1_MYCGA
23	51	48.1	1397	2	Q73P73_TREDE
24	51	48.1	1554	2	Q8KZL5_9STRE
25	50	47.2	384	2	Q8STT7_ENCCU
26	50	47.2	836	2	Q4TPT1_9SPHN
27	50	47.2	1122	1	ADP2_MYCGA
28	50	47.2	1290	2	Q48756_LEUME
29	50	47.2	1577	2	Q55265_STRSL
30	49	46.2	800	2	Q6BMJ4_DEBHA
31	49	46.2	958	2	Q7UWE0_RHOBA

32	49	46.2	2835	2	Q8G9Q2_LEUME
33	48.5	45.8	357	2	Q4UAA7_THEAN
34	48	45.3	413	2	Q7P7K6_FUSNV
35	48	45.3	493	2	Q75LW1_ORYSA
36	48	45.3	1365	1	GTF5_STRDO
37	48	45.3	1372	2	Q727C7_DESVH
38	48	45.3	1463	2	Q5SBM6_LACPE
39	48	45.3	2763	2	Q6FRZ9_CANGA
40	47	44.3	294	2	Q7Q263_GIALA
41	47	44.3	412	2	Q4SMC7_TETNG
42	47	44.3	436	2	Q7UQN6_RHOBA
43	47	44.3	553	2	Q7ZWT6_XENLA
44	47	44.3	567	2	Q4KHW9_PSEF5
45	47	44.3	575	2	Q7PMV0_ANOGA

ALIGNMENTS

RESULT 1  
GTF1\_STRDO  
ID GTF1\_STRDO STANDARD; PRT; 1597 AA.  
AC P11001;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 13-SEP-2005 (Rel. 48, Last annotation update)  
DE Glucosyltransferase-1 precursor (EC 2.4.1.5) (GTF-1) (Dextranucrase)  
DE (Sucrose 6-glucosyltransferase).  
GN Names=gtf1;  
OS Streptococcus downei (Streptococcus sobrinus).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1317;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
RC STRAIN=MFE28;  
RX MEDLINE=97308014; PubMed=3040686;  
RA Ferretti J.O., Gilpin M.L., Russell R.R.B.;  
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus sobrinus MFE28.";  
RL J. Bacteriol. 169:4271-4278(1987).  
CC -!- FUNCTION: Production of extracellular glucans, that are thought to play a key role in the development of the dental plaque because of their ability to adhere to smooth surfaces and mediate the aggregation of bacterial cells and food debris.  
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-fructose + (1,6-alpha-D-glucosyl)(n+1).  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- MISCELLANEOUS: GTF-1 synthesizes water-insoluble glucans (alpha 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes water-soluble glucans (alpha 1,6-glucose). GTF-Sr synthesizes both forms of glucans.  
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.  
CC -!- SIMILARITY: Contains 19 cell wall binding repeats.

-----  
This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.  
-----  
EMBL; M17391; AAC63063.1; -; Genomic DNA.  
DR InterPro; IPR002479; Cell wall bd put.  
DR InterPro; IPR003318; Glyco hydro\_70.  
DR Pfam; PF01473; CW binding\_I\_4.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
DR Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.  
FT SIGNAL 1 38 Potential.  
FT CHAIN 39 1597 Glucosyltransferase-I.  
FT REPEAT 1099 1132 A repeat.  
FT REPEAT 1163 1213 AC repeat.  
FT REPEAT 1227 1277 AC repeat.  
FT REPEAT 1292 1342 AC repeat.  
-----

```

FT REPEAT      1352 1399      B repeat.
FT REPEAT      1406 1455      AC repeat.
FT REPEAT      1465 1512      B repeat.
FT REPEAT      1519 1568      AC repeat.
FT REPEAT      1582 1597      A repeat (incomplete).
FT REGION      39 1050      Catalytic (approximate).
FT REGION      1099 1597      1.25 A, 2 B and 5 AC repeats.
FT REGION      1099 1597      Glucan-binding (approximate).
SQ SEQUENCE    1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match      100.0%; Score 106; DB 1; Length 1597;
Best Local Similarity 100.0%; Pred. No. 6.4e-08;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DGKLYRYDANSQDQAFNKS 20
Db      1444 DGKLYRYDANSQDQAFNKS 1463

RESULT 2
Q55263_9STRE
ID Q55263_9STRE PRELIMINARY; PRT; 1590 AA.
AC Q55263;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE GTF-I.
GN Name=Glucosyltransferase;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33478;
SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match      97.2%; Score 103; DB 2; Length 1590;
Best Local Similarity 95.0%; Pred. No. 2e-07;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 DGKLYRYDANSQDQAFNKS 20
Db      1438 DGKLYRYDANSQDQAFNKS 1457

RESULT 3
Q59983_9STRE
ID Q59983_9STRE PRELIMINARY; PRT; 1590 AA.
AC Q59983;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)

```

---

```

DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name=gtfI;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OMZ176;
RC MEDLINE=94146405; PubMed=8312602;
RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;
RA "DNA sequence of the glucosyltransferase gene of serotype d
RT Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
RN [2];
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSSP; P06653; 1HCX.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding 1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Signal; Transferase.
FT SIGNAL      1 38      Potential.
FT CHAIN        39 1590      Glucosyltransferase-I.
SQ SEQUENCE    1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match      97.2%; Score 103; DB 2; Length 1590;
Best Local Similarity 95.0%; Pred. No. 2e-07;
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 DGKLYRYDANSQDQAFNKS 20
Db      1438 DGKLYRYDANSQDQAFNKS 1457

RESULT 4
GTF2_STRDO
ID GTF2_STRDO STANDARD; PRT; 1592 AA.
AC P27470;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1];
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA]..
RC STRAIN=6715 / Serotype G;
RC MEDLINE=91123227; PubMed=1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
RA Kagawa H.;
RT "Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL J. Bacteriol. 173:989-996(1991).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.

```

KW Glycosyltransferase; Transferase..

SQ SEQUENCE 1454 AA; 160569 MW; 2DC7342963B6FD87 CRC64;

Query Match 67.0%; Score 71; DB 2; Length 1454;  
 Best Local Similarity 66.7%; Pred. No. 0.031;  
 Matches 12; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 DGKLRYYDANGSDQAFNK 18  
 |||:||||:|||||:  
 Db 1373 DGKMRYYDADSGDMVTNR 1390

RESULT 6

Q8VUH3\_STRMU  
 ID Q8VUH3\_STRMU PRELIMINARY; PRT; 591 AA.  
 AC Q8VUH3;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Glucosyltransferase SI (Fragment).  
 GN Name=gtfC;  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=XG;  
 RX MEDLINE=22322694; PubMed=12435673;  
 RX DOI=10.1128/AAC.46.12.3756-3764.2002;  
 RA Tsuda H., Yamashita Y., Shibata Y., Nakano Y., Koga T.;  
 RT "Genes involved in bacitracin resistance in Streptococcus mutans.";  
 RL Antimicrob. Agents Chemother. 46:3756-3764 (2002).  
 RL EMBL; AB078507; BAB83942.1; -; Genomic\_DNA.  
 DR HSSP; P07762; 1M7X.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW\_binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 2.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Transferase.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 591 AA; 67094 MW; 0933DCE4421DAF30 CRC64;

Query Match 61.3%; Score 65; DB 2; Length 591;  
 Best Local Similarity 60.0%; Pred. No. 0.11;  
 Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 DGKLRYYDANGSDQAFNKSV 20  
 |||||:||||:||||:  
 Db 363 DGKLYFDRDSGNQISNRFV 382

RESULT 7

GTFC\_STRMU  
 ID GTFC\_STRMU STANDARD; PRT; 1455 AA.  
 AC P13470; O69382; O69385; O69398; O69397; P05427;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-MAY-2005 (Rel. 47, Last annotation update)  
 DE Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI)  
 DE (dextrantransferase) (Sucrose 6-glucosyltransferase).  
 GN Name=gtfC; OrderedLocusNames=SMU.1005;  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=GS-5;  
 RX MEDLINE=89137980; PubMed=2976010; DOI=10.1016/0378-1119(88)90382-4;  
 RA S., Shiroza T., Kuramitsu H.K.;

"Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";  
Gene 69:101-109(1988).  
[2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,  
RC MT4467 / Serotype e, and MT8148 / Serotype C;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RT Streptococcus mutans";  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=UAI59 / ATCC 700610 / Serotype C;  
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;  
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,  
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,  
RA Ferretti J.J.;  
RT "Genome sequence of Streptococcus mutans UAI59, a cariogenic dental  
RT pathogen.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
RN [4]  
RP NUCLEOTIDE SEQUENCE OF 1-349.  
RC STRAIN=GS-5;  
RX MEDLINE=87308013; PubMed=3040685;  
RA Shiroza T., Ueda S., Kuramitsu H.K.;  
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";  
RL J. Bacteriol. 169:4263-4270(1987).  
CC -1- FUNCTION: Production of extracellular glucans, that are thought to  
CC play a key role in the development of the dental plaque because of  
CC their ability to adhere to smooth surfaces and mediate the  
CC aggregation of bacterial cells and food debris.  
CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-  
CC fructose + (1,6-alpha-D-glucosyl)(n+1).  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha  
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes both  
CC water-soluble glucans (alpha 1,6-glucose), GTF-SI synthesizes both  
CC forms of glucans.  
CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.  
CC -1- SIMILARITY: Contains 5 cell wall binding repeats.  
CC -----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.  
CC -----  
DR EMBL; M22054; AAA88592.1; -; Genomic DNA.  
DR EMBL; D88652; BAA26102.1; -; Genomic DNA.  
DR EMBL; D88653; BAA26106.1; -; Genomic DNA.  
DR EMBL; D88658; BAA26110.1; -; Genomic DNA.  
DR EMBL; D88661; BAA26114.1; -; Genomic DNA.  
DR EMBL; D89978; BAA26120.1; -; Genomic DNA.  
DR EMBL; AE014940; AAN58706.1; -; Genomic DNA.  
DR EMBL; M17361; AAA88589.1; -; Genomic DNA.  
DR PIR; JTO345; JTO345.  
DR HSSP; P06653; IH8G.  
DR InterPro; IPR002479; Cell wall bd put.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 2.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
DR Transferrase.  
KW Complete proteome; Dental caries; Glycosyltransferase; Repeat; Signal;  
FT FT  
FT CHAIN 35 1455 Glucosyltransferase-SI.  
FT REPEAT 1126 1159 A repeat.  
FT REPEAT 1169 1200 A repeat.  
FT REPEAT 1227 1238 A repeat.  
FT REPEAT 1253 1303 AC repeat.  
FT REPEAT 1318 1330 A repeat (incomplete).  
FT

FT REGION 35 1050 Catalytic (approximate).  
FT REGION 1126 1455 2,4 A, 1 C and 1 AC repeats.  
FT REGION 1126 1455 Glucan-binding (approximate).  
FT VARIANT 21 21 V -> I (in strain GS-5).  
FT VARIANT 81 81 P -> L (in strain MT4239).  
FT VARIANT 106 106 D -> V (in strain GS-5).  
FT VARIANT 116 116 S -> A (in strain GS-5 and strain  
MT4467).  
FT VARIANT 126 126 A -> T (in strain GS-5).  
FT VARIANT 150 151 SR -> PK (in strain GS-5, strain MT4239  
and strain MT4467).  
FT VARIANT 256 256 A -> V (in strain GS-5 and strain  
MT4467).  
FT VARIANT 425 425 R -> N (in strain MT4251).  
FT VARIANT 519 519 Y -> D (in strain MT4245 and strain  
MT4251).  
FT VARIANT 538 538 R -> K (in strain MT4245 and strain  
MT4251).  
FT VARIANT 545 545 Y -> F (in strain MT4245 and strain  
MT4251).  
FT VARIANT 597 597 N -> D (in strain MT4245, strain MT4251,  
strain MT4467 and strain MT8148).  
FT VARIANT 600 600 R -> K (in strain MT4245, strain MT4251,  
strain MT4467 and strain MT8148).  
FT VARIANT 601 601 A -> T (in strain GS-5).  
FT VARIANT 614 614 M -> T (in strain GS-5).  
FT VARIANT 727 727 T -> I (in strain MT8148).  
FT VARIANT 734 734 A -> V (in strain MT8148).  
FT VARIANT 964 964 L -> F (in strain MT4239).  
FT VARIANT 1113 1113 N -> Y (in strain MT4239).  
FT VARIANT 1118 1118 I -> T (in strain MT4239).  
FT VARIANT 1204 1204 I -> V (in strain GS-5, strain MT4239,  
strain MT4467 and strain MT8148).  
FT VARIANT 1208 1208 V -> I (in strain MT8148).  
FT VARIANT 1292 1294 DGH -> NGY (in strain GS-5, strain MT4467  
and strain MT8148).  
FT VARIANT 1305 1369 Missing (in strain MT4245).  
FT VARIANT 1326 1326 I -> V (in strain MT8148).  
FT VARIANT 1331 1331 T -> A (in strain GS-5, strain MT4239,  
strain MT4467 and strain MT8148).  
FT VARIANT 1377 1377 R -> K (in strain MT8148).  
FT VARIANT 1398 1398 V -> I (in strain MT8148).  
FT VARIANT 1424 1424 V -> N (in strain MT4239).  
FT VARIANT 1439 1439 V -> I (in strain MT4239 and strain  
MT8148).  
FT VARIANT 1444 1444 S -> P (in strain MT8148).  
FT CONFLICT 1337 1455 QRLYFKNSGVQAKGELITERKRIKIKYDPNSGNEVNRNRYVR  
TSGNMYFYGNDGYALIGHWVVEGRVYFDENGVIYASHD  
QRNWDYDVRDRDFRGSSAVRPRSRNGFPDNFFRF ->  
HASILSLMVFRLRESSLSQSVKVSNTLIPENKFKVIVM  
(in Ref. 1).  
SQ SEQUENCE 1455 AA; 162966 MW; 3CB455A9A4FEC86 CRC64;  
Query Match 61.3%; Score 65; DB 1; Length 1455;  
Best Local Similarity 60.0%; Pred. No. 0.29;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 1 DGKLYRDANSQDAFNKSV 20  
| | | | | : | | : | |  
DB 1227 DGKLYRDRDSGNQISNRFV 1246  
RESULT 8  
GTFD\_STRMU  
ID GTFD\_STRMU STANDARD; PRT; 1462 AA.  
AC P49331; O69386; O69389; O69392; O69398;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-MAY-2005 (Rel. 47, Last annotation update)  
DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)  
DE (Sucrose 6-glucosyltransferase).  
GN Name=gtfD; OrderedLocustNames=SMU.910;  
OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=GS-5;  
 RX MEDLINE=91100958; PubMed=2148600;  
 RA Honda O., Kato C., Kuramitsu H.K.;  
 RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
 the glucosyltransferase-S enzyme.";  
 RL J. Gen. Microbiol. 136:2099-2105(1990).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,  
 RC MT4467 / Serotype e, and MT8148 / Serotype c;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 RA Kimura S., Hamada S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 Streptococcus mutans.";  
 RL FEMS Microbiol. Lett. 161:331-336(1998).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=UA159 / ATCC 700610 / Serotype c;  
 RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;  
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,  
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,  
 RA Ferretti J.J.;  
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental  
 pathogen.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
 CC -!- FUNCTION: Production of extracellular glucans, that are thought to  
 play a key role in the development of the dental plaque because of  
 their ability to adhere to smooth surfaces and mediate the  
 aggregation of bacterial cells and food debris.  
 CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-  
 fructose + (1,6-alpha-D-glucosyl)(n+1).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha  
 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes  
 water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both  
 forms of glucans.  
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.  
 CC -!- SIMILARITY: Contains 6 cell wall binding repeats.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use as long as its content is in no way modified and this statement is not  
 removed.  
 CC -----  
 CC EMBL; M29296; AAA26895.1; -; Genomic DNA.  
 CC EMBL; D88653; BAA26103.1; -; Genomic DNA.  
 CC EMBL; D88656; BAA26107.1; -; Genomic DNA.  
 CC EMBL; D88659; BAA26111.1; -; Genomic DNA.  
 CC EMBL; D88662; BAA26115.1; -; Genomic DNA.  
 CC EMBL; D89799; BAA26121.1; -; Genomic DNA.  
 CC EMBL; A5014932; AAN58619.1; -; Genomic DNA.  
 CC HSSP; P06653; IGVM.  
 DR InterPro; IPR002479; Cell\_wall\_bd\_put.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 4.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;  
 KW Transferase.  
 FT SIGNAL 1 ? Potential.  
 FT CHAIN ? 1462 Glucosyltransferase-S.  
 FT REPEAT 1232 1295 1.  
 FT REPEAT 1296 1359 2.  
 FT REPEAT 1360 1423 3.  
 FT REGION 1232 1423 3 X 63 AA approximate tandem repeats.  
 FT VARIANT 10 10 Y -> H (in strain GS-5, strain MT4239,

FT strain MT4245, strain MT4251, strain  
 FT MT4467 and strain MT8148).  
 FT I -> V (in strain GS-5, strain MT4239,  
 FT strain MT4245, strain MT4251, strain  
 FT MT4467 and strain MT8148).  
 FT K -> E (in strain MT4467).  
 FT A -> S (in strain MT4239 and strain  
 FT MT4245).  
 FT A -> T (in strain MT4251 and strain  
 FT MT8148).  
 FT T -> I (in strain MT4239 and strain  
 FT MT4245).  
 FT A -> V (in strain MT4239, strain MT4245  
 FT and strain MT8148).  
 FT A -> S (in strain GS-5 and strain  
 FT MT4467).  
 FT A -> V (in strain MT4245).  
 FT A -> T (in strain GS-5, strain MT4239,  
 FT strain MT4245, strain MT4251, strain  
 FT MT4467 and strain MT8148).  
 FT V -> L (in strain MT4239).  
 FT D -> N (in strain MT8148).  
 FT E -> D (in strain MT4239, strain MT4245  
 FT and strain MT4251).  
 FT D -> N (in strain MT4239, strain MT4245  
 FT and strain MT4251).  
 FT Q -> H (in strain MT4245).  
 FT D -> N (in strain MT4239 and strain  
 FT MT4251).  
 FT E -> K (in strain MT4239).  
 FT V -> F (in strain MT4239).  
 FT F -> L (in strain MT4239, strain MT4251  
 FT and strain MT4467).  
 FT KKKYQTQ -> EKEYTL (in strain MT4251).  
 FT A -> S (in strain MT4239).  
 FT TQOGSEA -> ADKGND (in strain MT4251).  
 FT TQGS -> ADKGN (in strain MT4239 and  
 FT strain MT4245).  
 FT T -> A (in strain GS-5, strain MT4239,  
 FT strain MT4245, strain MT4251, strain  
 FT MT4467 and strain MT8148).  
 FT D -> Y (in strain MT4251).  
 FT E -> K (in strain MT4245 and strain  
 FT MT4251).  
 FT LG -> IR (in strain MT4251).  
 FT G -> R (in strain MT4245).  
 FT G -> R (in strain MT4239).  
 FT H -> Q (in strain GS-5).  
 FT S -> N (in strain MT4239).  
 FT Y -> C (in strain MT4251 and strain  
 FT MT4467).  
 FT F -> L (in strain MT4467).  
 FT Q -> P (in strain MT4245).  
 FT K -> T (in strain MT4245).  
 FT N -> D (in strain MT4245).  
 FT G -> D (in strain GS-5 and strain  
 FT MT4467).  
 FT G -> R (in strain GS-5).  
 FT R -> K (in strain MT4467).  
 FT RYDKNSGNMVKVYTLANGRIGIDRWGIARYY -> VY  
 FT R (in Ref. 1).  
 SQ SEQUENCE 1462 AA; 16338 MW; CB4A279C4D708645 CRC64;  
 Query Match 61.1%; Score 65; DB 1; Length 1462;  
 Best Local Similarity 63.2%; Pred. No. 0.3;  
 .Matches 12; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 2 GKRLRYDANSQGFQAFKSV 20  
 |||||  
 Db 1425 GLLRYDKSGNMVYKVV 1443

RESULT 9

```

O68542_STRSL
ID O68542_STRSL PRELIMINARY; PRT; 1449 AA.
AC O68542;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase N (Fragment).
GN Name=gtfN;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=VI477;
RA Jaffe R.I.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR ENBL; AF049609; AAC05156.1; -; Genomic_DNA.
DR PIR; T30552; T30552.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
FT NON TER.
SQ SEQUENCE 1449 AA; 159895 MW; 0700F6D748471BFB CRC64;

Query Match 60.4%; Score 64; DB 2; Length 1449;
Best Local Similarity 61.1%; Pred. No. 0.43;
Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 DGKLRYYDANSQDAFNK 18
DB 1285 NGAIRYYDANSGEWARNR 1302

RESULT 10
Q5264_STRSL
ID Q5264_STRSL PRELIMINARY; PRT; 1449 AA.
AC Q5264;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase precursor.
GN Name=gtfL;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=95122197; PubMed=7822030;
RA Simpson C.L., Giffard P.M., Jacques N.A.;
RT "Streptococcus salivarius ATCC 25975 possesses at least two genes
RT coding for primer-independent glucosyltransferases.";
RL Infect. Immun. 63:609-621(1995).
DR ENBL; L35495; AAC41412.1; -; Genomic_DNA.
DR PIR; T30857; T30857.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Signal; Transferase.
FT SIGNAL 1 35 Potential.
FT CHAIN 35 1449 glucosyltransferase.
SQ SEQUENCE 1449 AA; 159984 MW; DD62F07306E86A46 CRC64;

Query Match 60.4%; Score 64; DB 2; Length 1449;
Best Local Similarity 61.1%; Pred. No. 0.43;

```

```

Matches 11; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 DGKLRYYDANSQDAFNK 18
DB 1285 NGAIRYYDANSGEWARNR 1302

RESULT 11
Q9LCH3_STROR
ID Q9LCH3_STROR PRELIMINARY; PRT; 1575 AA.
AC Q9LCH3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Glucosyltransferase.
GN Name=gtfR;
OS Streptococcus oralis.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1303;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC10557;
RX MEDLINE=20231779; PubMed=10768934;
RX DOI=10.1128/IAI.68.5.2475-2483.2000;
RA Fujiwara T., Hoshino T., Ooshima T., Sobue S., Hamada S.;
RT "Purification, characterization, and molecular analysis of the gene
RT encoding glucosyltransferase from Streptococcus oralis.";
RL Infect. Immun. 68:2475-2483(2000).
DR ENBL; AB025228; BAA95201.1; -; Genomic_DNA.
DR HSSP; P06653; 1HCX.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1575 AA; 176792 MW; 772A26E4D7C2E543 CRC64;

Query Match 60.4%; Score 64; DB 2; Length 1575;
Best Local Similarity 60.0%; Pred. No. 0.47;
Matches 12; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 DGKLRYYDANSQDAFNKSV 20
DB 1470 DKSIRYFDANSGEWATNKFV 1489

RESULT 12
GTFB_STRMU
ID GTFB_STRMU STANDARD; PRT; 1476 AA.
AC P08987; O69381; O69384; O69387; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfB; OrderedLocusNames=SMU.1004;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RL J. Bacteriol. 169:4263-4270(1987).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,

```

RC MT4467 / Serotype e, and MT8148 / Serotype c;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RL Streptococcus mutans";  
RN FEMS Microbiol. Lett. 161:331-336(1998).  
RN [3].  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=UA159 / ATCC 700610 / Serotype c;  
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;  
RA Aidic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,  
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lal H., White J., Roe B.A.,  
RA Ferretti J.J.;  
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental  
RT pathogen";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
CC -1- FUNCTION: Production of extracellular glucans, that are thought to  
CC play a key role in the development of the dental plaque because of  
CC their ability to adhere to smooth surfaces and mediate the  
CC aggregation of bacterial cells and food debris.  
CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-  
CC fructose + (1,6-alpha-D-glucosyl)(n+1).  
CC -1- SUBCELLULAR LOCATION: Secreted.  
CC -1- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha  
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes  
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both  
CC forms of glucans.  
CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.  
CC -1- SIMILARITY: Contains 10 cell wall binding repeats.  
CC -----  
CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use as long as its content is in no way modified and this statement is not  
CC removed.  
CC -----  
DR EMBL; M17361; AAA88588.1; -; Genomic DNA.  
DR EMBL; D88651; BAA26101.1; -; Genomic DNA.  
DR EMBL; D88654; BAA26105.1; -; Genomic DNA.  
DR EMBL; D88657; BAA26109.1; -; Genomic DNA.  
DR EMBL; D88660; BAA26113.1; -; Genomic DNA.  
DR EMBL; D89977; BAA26119.1; -; Genomic DNA.  
DR EMBL; AE014940; AAN58705.1; -; Genomic DNA.  
DR PIR; B33135; B33135.  
DR HSSP; P06653; 1H8G.  
DR InterPro; IPR002479; Cell\_wall\_bd\_put.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 4.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
DR Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;  
KW Transferrase.  
FT SIGNAL 1 34 Potential.  
FT CHAIN 35 1476 Glucosyltransferase-I.  
FT REPEAT 1097 1130 A repeat.  
FT REPEAT 1161 1210 1.  
FT REPEAT 1225 1275 2.  
FT REPEAT 1290 1340 3.  
FT REPEAT 1355 1405 4.  
FT REPEAT 1420 1470 5.  
FT REGION 1097 1476 Catalytic (approximate).  
FT REGION 1161 1470 Glucan-binding (approximate).  
FT REGION 1161 1470 5 X tandem repeats.  
FT VARIANT 62 62 S -> T (in strain MT4239).  
FT VARIANT 65 65 T -> I (in strain GS-5).  
FT VARIANT 68 68 V -> A (in strain GS-5, strain MT4245,  
FT strain MT4251, strain MT4467 and strain  
FT MT8148).  
FT VARIANT 78 78 Q -> P (in strain MT4251).  
FT VARIANT 86 86 I -> S (in strain GS-5, strain MT4245,  
FT strain MT4251, strain MT4467 and strain  
FT MT8148).  
FT

FT VARIANT 89 89 S -> F (in strain MT4251).  
FT VARIANT 168 168 K -> D (in strain GS-5, strain MT4467 and  
FT strain MT8148).  
FT VARIANT 276 276 N -> R (in strain MT4239).  
FT VARIANT 399 399 I -> T (in strain MT4239).  
FT VARIANT 474 474 K -> R (in strain MT8148).  
FT VARIANT 512 512 F -> Y (in strain MT8148).  
FT VARIANT 519 519 T -> I (in strain MT8148).  
FT VARIANT 701 701 A -> V (in strain MT8148).  
FT VARIANT 708 708 F -> L (in strain MT8148).  
FT VARIANT 938 938 FGKPE -> VGTPEVA (in strain GS-5, strain  
FT MT4239 and strain MT4467).  
FT VARIANT 952 957 SV -> NT (in strain GS-5, strain MT4239  
FT and strain MT4467).  
FT VARIANT 963 964 ADS -> VDQ (in strain GS-5, strain MT4239  
FT and strain MT4467).  
FT VARIANT 968 970 A -> T (in strain MT4239).  
FT VARIANT 1086 1086 S -> N (in strain MT4239).  
FT VARIANT 1158 1158 H -> Y (in strain MT4251).  
FT VARIANT 1163 1163 E -> K (in strain MT8148).  
FT VARIANT 1168 1168 Y -> C (in strain MT8148).  
FT VARIANT 1182 1182 R -> P (in strain MT4239).  
FT VARIANT 1234 1234 A -> H (in strain GS-5 and strain  
FT MT4467).  
FT VARIANT 1263 1263 R -> P (in strain MT8148).  
FT VARIANT 1263 1263 Y -> H (in strain GS-5, strain MT4239,  
FT strain MT4467 and strain MT8148).  
FT VARIANT 1264 1264 S -> G (in strain GS-5, strain MT4239,  
FT strain MT4467 and strain MT8148).  
FT VARIANT 1272 1272 H -> Y (in strain GS-5 and strain  
FT MT4467).  
FT VARIANT 1329 1329 Y -> H (in strain GS-5, strain MT4239,  
FT strain MT4467 and strain MT8148).  
FT VARIANT 1394 1394 S -> G (in strain GS-5, strain MT4239,  
FT strain MT4467 and strain MT8148).  
FT VARIANT 1402 1402 Y -> H (in strain MT4467).  
FT VARIANT 1459 1459 R -> A (in Ref. 1).  
FT CONFLICT 570 570 ADQDVRVAASTAPSTDGK -> LIKMFALRLARPHQQA  
FT CONFLICT 800 817 (in Ref. 1).  
FT CONFLICT 1310 1310 H -> L (in Ref. 1).  
SQ SEQUENCE 1476 AA; 165847 MW; 9C6E09F731B4CBCF CRC64;  
Query Match 59.4%; Score 63; DB 1; Length 1476;  
Best Local Similarity 57.9%; Pred. No. 0.64;  
Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;  
QY 2 GKLYRYDANSQDQAFNKSV 20  
Db 1265 GRISYDYSNSGDIQNRNFV 1283  
RESULT 13  
Q54178\_STRGN  
ID Q54178\_STRGN PRELIMINARY; PRT; 1577 AA.  
AC Q54178; Q54247;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Glucosyltransferase.  
GN Name=gtfG;  
OS Streptococcus gordonii.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1302;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=Challis;  
RX MEDLINE=96157084; PubMed=8586195;  
RA Vickerman M.M., Sulavik M.C., Clewell D.B.;  
RT "Molecular analysis of Streptococcus gordonii glucosyltransferase  
RT phase variants.";  
RL Dev. Biol. Stand. 85:309-314(1995).  
FT

```
RN [2]
RC NUCLEOTIDE SEQUENCE.
RA Minick P., Vickerman M.;
RL SUBMITTED (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U12643; AAC43483.1; -; Genomic_DNA.
DR PIR; B41898; B41898.
DR HSP; P06653; 1HCX.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 5.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase
SQ SEQUENCE 1577 AA; 177805 MW; 5AE0328DC5E08D18 CRC64;

Query Match 59.4%; Score 63; DB 2; Length 1577;
Best Local Similarity 61.1%; Pred. No. 0.69;
Matches 11; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 DGKLYDYDANSQDAFNK 18
Db 1342 DKSIRYFDANSGEAVNK 1359

RESULT 14
Q8DUW9_STRMU
ID Q8DUW9_STRMU PRELIMINARY; PRT; 726 AA.
AC Q8DUW9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Glucan-binding protein D with lipase activity; BglB-like protein.
GN Name=gbbp; OrderedLocusNames=SMU.772;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RC STRAIN=UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lal H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR EMBL; AE014918; AAN58492.1; -; Genomic_DNA.
DR HSP; P06653; 1H8G.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR005877; GpOs_YsIRK.
DR InterPro; IPR000379; Ser_estrs.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF04650; YsIRK_signal; 1.
DR TIGRFAMs; TIGR01168; YsIRK_signal; 1.
KW Complete proteome.
SQ SEQUENCE 726 AA; 79786 MW; B2AB7C06F6E7CB4E CRC64;

Query Match 58.5%; Score 62; DB 2; Length 726;
Best Local Similarity 84.6%; Pred. No. 0.41;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DGKLYDYDANSQD 13
Db 262 DGSILRYDYDKNSGD 274

RESULT 15
```

```
TOLB WOLSU
ID TOLB WOLSU STANDARD; PRT; 421 AA.
AC Q7MA16;
DT 13-SEP-2005 (Rel. 48, Created)
DT 13-SEP-2005 (Rel. 48, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE TolB protein precursor.
GN Name=tolB; OrderedLocusNames=WS0521;
OS Wolinella succinogenes.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Helicobacteraceae; Wolinella.
OX NCBI_TaxID=844;
RN [1]
RC NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX MEDLINE=22882897; PubMed=14500908; DOI=10.1073/pnas.1932838100;
RA Baar C., Eppinger M., Raddatz G., Simon J., Lanz C., Klumpek O.,
RA Nandakumar R., Gross R., Rosinus A., Keller H., Jagtap P., Linke B.,
RA Meyer F., Lederer H., Schuster S.C.;
RT "Complete genome sequence and analysis of Wolinella succinogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:11690-11695(2003).
CC -!- FUNCTION: Involved in the tonB-independent uptake of proteins (By
similarity).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to the tolB family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
CC -----
DR EMBL; BX571658; CAB09658.1; -; Genomic_DNA.
DR HAMAP; MF_00671; -; 1.
DR InterPro; IPR007195; TolB_N.
DR Pfam; PF04052; TolB_N; 1.
KW Complete proteome; Periplasmic; Protein transport; Signal; Transport.
FT SIGNAL 1 16 Potential.
FT CHAIN 17 421 TolB protein.
SQ SEQUENCE 421 AA; 47680 MW; 8749130785BE8E0D CRC64;

Query Match 53.8%; Score 57; DB 1; Length 421;
Best Local Similarity 64.7%; Pred. No. 1.5;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 3 KLYRYDANSQDAFNKS 19
Db 107 ELRLYDVSNGDSVFAKS 123

Search completed: February 10, 2006, 22:31:16
Job time : 15.7298 secs
```



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:31:39 ; Search time 3.60887 Seconds  
(without alignments)  
458.180 Million cell updates/sec

Title: US-10-797-821-27  
Perfect score: 106  
Sequence: 1 DGKLRVDANSQDAFNKSV 20

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/PCUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	106	100.0	20	1	US-08-057-162B-4
2	65	61.3	1375	2	US-09-210-361-4
3	65	61.3	1375	2	US-09-740-274-4
4	62	58.5	1475	2	US-09-007-999-2
5	62	58.5	1475	2	US-09-210-361-2
6	62	58.5	1475	2	US-09-740-274-2
7	56	52.8	1430	2	US-09-008-172-2
8	56	52.8	1430	2	US-09-210-361-6
9	56	52.8	1430	2	US-09-740-274-6
10	50	47.2	1577	1	US-08-793-824-2
11	47	44.3	1256	2	US-09-248-796A-18057
12	45	42.5	321	2	US-09-107-532A-4138
13	44.5	42.0	62	2	US-09-248-796A-24375
14	44	41.5	393	2	US-09-328-352-7939
15	44	41.5	859	2	US-09-538-092-206
16	43	40.6	116	2	US-09-377-285B-69
17	43	40.6	116	2	US-10-192-381-69
18	43	40.6	194	2	US-09-605-703B-1576
19	43	40.6	198	2	US-09-248-796A-19506
20	43	40.6	246	2	US-09-902-540-13920
21	43	40.6	378	1	US-08-986-217-6
22	43	40.6	380	1	US-09-026-587-4
23	43	40.6	380	1	US-09-237-420-4
24	43	40.6	380	2	US-09-387-811-4
25	43	40.6	380	2	US-09-823-240A-10
26	43	40.6	381	2	US-09-949-016-7291
27	43	40.6	635	2	US-09-417-197-125

28	43	40.6	740	2	US-09-902-540-11267	Sequence 11267, A
29	42.5	40.1	786	2	US-09-252-991A-32352	Sequence 32352, A
30	42	39.6	236	2	US-09-248-796A-23590	Sequence 23590, A
31	42	39.6	281	2	US-09-949-016-6831	Sequence 6831, Ap
32	42	39.6	324	2	US-09-949-016-7870	Sequence 7870, Ap
33	42	39.6	411	1	US-08-440-845D-3	Sequence 3, Appli
34	42	39.6	411	2	US-08-868-458-3	Sequence 3, Appli
35	42	39.6	772	1	US-08-714-677-2	Sequence 2, Appli
36	42	39.6	772	1	US-08-393-540-2	Sequence 2, Appli
37	42	39.6	772	1	US-08-714-537-2	Sequence 4, Appli
38	42	39.6	775	1	US-08-714-677-4	Sequence 4, Appli
39	42	39.6	775	1	US-08-393-540-4	Sequence 4, Appli
40	42	39.6	775	1	US-08-714-537-4	Sequence 4, Appli
41	42	39.6	1237	2	US-09-332-522E-6	Sequence 6, Appli
42	42	39.6	1262	2	US-09-198-452A-97	Sequence 97, Appli
43	42	39.6	1266	2	US-09-438-185A-83	Sequence 83, Appli
44	41.5	39.2	421	2	US-09-248-796A-17773	Sequence 17773, A
45	41	38.7	94	2	US-09-270-767-31984	Sequence 31984, A

ALIGNMENTS

RESULT 1  
US-08-057-162B-4  
; Sequence 4, Application US/08057162B  
; Patent No. 5686075  
; GENERAL INFORMATION:  
; APPLICANT: Taubman, Martin A.  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR DENTAL CARIES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/057,162B  
; FILING DATE: 30-APR-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/877,295  
; FILING DATE: 01-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wagner, Richard W.  
; REGISTRATION NUMBER: 34,480  
; REFERENCE/DOCKET NUMBER: FDC92-01A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-057-162B-4

Query Match 100.0%; Score 106; DB 1; Length 20;  
Best Local Similarity 100.0%; Pred. No. 8.8e-11;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGKLRVDANSQDAFNKSV 20  
|||||



```
Best Local Similarity 57.9%; Pred. No. 0.2;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 GKLYRYDANSQDAFNKSV 20
Db 1264 GRISYYDGNSGDQIRNFV 1282

RESULT 6
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; SOFTWARE: FastSeq for Windows Version 3.0
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 1475
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 58.5%; Score 62; DB 2; Length 1475;
Best Local Similarity 57.9%; Pred. No. 0.2;
Matches 11; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 2 GKLYRYDANSQDAFNKSV 20
Db 1264 GRISYYDGNSGDQIRNFV 1282

RESULT 7
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; CURRENT FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match 52.8%; Score 56; DB 2; Length 1430;
Best Local Similarity 55.6%; Pred. No. 1.9;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 DGKLYRYDANSQDAFNK 18
Db 1231 DGKKRYFDDGSGNMAVNR 1248

RESULT 8
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; SOFTWARE: FastSeq for Windows Version 3.0
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 6
; TYPE: PRT
; LENGTH: 1430
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match 52.8%; Score 56; DB 2; Length 1430;
Best Local Similarity 55.6%; Pred. No. 1.9;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 DGKLYRYDANSQDAFNK 18
Db 1231 DGKKRYFDDGSGNMAVNR 1248

RESULT 9
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; TYPE: PRT
; LENGTH: 1430
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match 52.8%; Score 56; DB 2; Length 1430;
Best Local Similarity 55.6%; Pred. No. 1.9;
Matches 10; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
```



```
; LOCATION: (8) LOCATION 1...321
; SEQUENCE DESCRIPTION: SEQ ID NO: 4138:
US-09-107-532A-4138

Query Match      42.5%; Score 45; DB 2; Length 321;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 10; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 1 DGKLYYDANSQDAFNKSV 20
Db 268 DGLKSMDFNSVDSFIQSV 287

RESULT 13
US-09-248-796A-24375
; Sequence 24375, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248.796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 24375
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-24375

Query Match      42.0%; Score 44.5; DB 2; Length 62;
Best Local Similarity 50.0%; Pred. No. 4.2;
Matches 10; Conservative 3; Mismatches 6; Indels 1; Gaps 1;

Qy 1 DGKLYYDANSQDAFNKSV 20
Db 41 EAKLYYDQPSISK-FNKRI 59

RESULT 14
US-09-328-352-7939
; Sequence 7939, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328.352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 7939
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-7939

Query Match      41.5%; Score 44; DB 2; Length 393;
Best Local Similarity 57.1%; Pred. No. 41;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 DGKLYYDANSQDQ 14
Db 304 DGKLTAYDLTTGEQ 317

RESULT 15
US-09-538-092-206
```

```
; Sequence 206, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538.092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratPatSeqFormatter Version 0.9
; SEQ ID NO 206
; LENGTH: 859
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number YDR416W
US-09-538-092-206

Query Match      41.5%; Score 44; DB 2; Length 859;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 8 DANSGDQAFNKS 19
Db 559 DANSSDKIFNKT 570

Search completed: February 10, 2006, 22:36:15
Job time : 3.60887 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:14:04 ; Search time 11.7339 Seconds  
(without alignments)  
712.176 Million cell updates/sec

Title: US-10-797-821-27

Perfect score: 106

Sequence: 1 DGKLRYYDANSQDAFNKSV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications\_AA\_Main.\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	106	100.0	20	4	US-10-383-930-27
2	106	100.0	20	5	US-10-797-821-27
3	103	97.2	1590	4	US-10-383-930-37
4	103	97.2	1590	5	US-10-797-821-37
5	65	61.3	1375	3	US-09-740-274-34
6	65	61.3	1375	4	US-10-383-930-35
7	65	61.3	1375	5	US-10-797-821-35
8	62	58.5	1475	4	US-10-282-122A-72033
9	62	58.5	1475	3	US-09-740-274-2
10	62	58.5	1475	4	US-10-383-930-34
11	62	58.5	1475	5	US-10-797-821-34
12	57	53.8	1497	5	US-10-484-218-18
13	56	52.8	1430	3	US-09-740-274-6
14	56	52.8	1430	4	US-10-383-930-36
15	56	52.8	1430	5	US-10-797-821-36
16	54	50.9	661	4	US-10-369-493-10446
17	53	50.0	1595	5	US-10-484-218-20
18	51	48.1	1554	4	US-10-383-930-38
19	51	48.1	1554	5	US-10-797-821-38
20	49	46.2	101	4	US-10-425-115-257044
21	48	45.3	493	4	US-10-437-963-169162
22	48	45.3	1006	5	US-10-484-218-22
23	48	45.3	1365	4	US-10-383-930-39
24	48	45.3	1365	5	US-10-797-821-39
25	47	44.3	695	4	US-10-369-493-22333
26	47	44.3	1448	4	US-10-032-585-7452
27	46	43.4	1112	5	US-10-450-763-52772

28	45	42.5	57	4	US-10-424-599-178427	Sequence 178427, A
29	45	42.5	73	4	US-10-767-701-51677	Sequence 51677, A
30	45	42.5	118	4	US-10-425-115-194703	Sequence 194703, A
31	45	42.5	340	4	US-10-767-701-47205	Sequence 47205, A
32	45	42.5	415	4	US-10-767-701-43900	Sequence 43900, A
33	45	42.5	419	4	US-10-425-114-61169	Sequence 61169, A
34	45	42.5	542	4	US-10-424-599-146181	Sequence 146181, A
35	45	42.5	717	4	US-10-369-493-2178	Sequence 2178, Ap
36	45	42.5	754	5	US-10-485-986-9	Sequence 9, Appli
37	44	41.5	228	4	US-10-425-114-71491	Sequence 71491, A
38	44	41.5	325	5	US-10-732-923-7451	Sequence 7451, Ap
39	44	41.5	381	4	US-10-282-122A-44675	Sequence 44675, A
40	44	41.5	478	4	US-10-425-115-240549	Sequence 240549, A
41	44	41.5	496	4	US-10-369-493-16572	Sequence 16572, A
42	44	41.5	729	4	US-10-282-122A-46259	Sequence 46259, A
43	44	41.5	1238	4	US-10-282-122A-52987	Sequence 52987, A
44	43	40.6	62	4	US-10-425-114-61187	Sequence 61187, A
45	43	40.6	116	4	US-10-192-381-69	Sequence 69, Appl

ALIGNMENTS

RESULT 1  
US-10-383-930-27  
; Sequence 27, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 27  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-27

Query Match 100.0%; Score 106; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGKLRYYDANSQDAFNKSV 20  
|||||  
Db 1 DGKLRYYDANSQDAFNKSV 20

RESULT 2

US-10-797-821-27  
; Sequence 27, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049

; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 27  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Glucan-Binding Domain GTF Peptide  
US-10-797-821-27

Query Match 100.0%; Score 106; DB 5; Length 20;  
Best Local Similarity 100.0%; Pred. No. 1.4e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGKLYRYDANSGDQAFNKS V 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 DGKLYRYDANSGDQAFNKS V 20

RESULT 3  
US-10-383-930-37  
; Sequence 37, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Taubman, Daniel J  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 37  
; LENGTH: 1590  
; TYPE: PRT  
; ORGANISM: Streptococcus sobrinus  
US-10-383-930-37

Query Match 97.2%; Score 103; DB 4; Length 1590;  
Best Local Similarity 95.0%; Pred. No. 4.4e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGKLYRYDANSGDQAFNKS V 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1438 DGKLYRYDANSGDQAFNKS V 1457

RESULT 4  
US-10-797-821-37  
; Sequence 37, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Taubman, Daniel J.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08

; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 37  
; LENGTH: 1590  
; TYPE: PRT  
; ORGANISM: Streptococcus sobrinus  
US-10-797-821-37

Query Match 97.2%; Score 103; DB 5; Length 1590;  
Best Local Similarity 95.0%; Pred. No. 4.4e-07;  
Matches 19; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGKLYRYDANSGDQAFNKS V 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1438 DGKLYRYDANSGDQAFNKS V 1457

RESULT 5  
US-09-740-274-4  
; Sequence 4, Application US/09740274  
; Patent No. US20020031826A1  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 61.3%; Score 65; DB 3; Length 1375;  
Best Local Similarity 60.0%; Pred. No. 0.36;  
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 DGKLYRYDANSGDQAFNKS V 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1227 DGKLYRYDANSGDQAFNKS V 1246

RESULT 6  
US-10-383-930-35  
; Sequence 35, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018



```

; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-35

Query Match      61.3%; Score 65; DB 4; Length 1375;
Best Local Similarity 60.0%; Pred. No. 0.36;
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 DGKLYYDANSQDAFNKSV 20
Db 1227 DGKLYFRDSDGNQISNRFV 1246

RESULT 7
US-10-797-821-35
; Sequence 35, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT FILING DATE: 2004-03-09
; CURRENT APPLICATION NUMBER: US/10/797,821
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 35
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-35

Query Match      61.3%; Score 65; DB 5; Length 1375;
Best Local Similarity 60.0%; Pred. No. 0.36;
Matches 12; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 1 DGKLYYDANSQDAFNKSV 20
Db 1227 DGKLYFRDSDGNQISNRFV 1246

RESULT 8
US-10-282-122A-72033
; Sequence 72033, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari

```

```

; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72033
; LENGTH: 726
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-282-122A-72033

Query Match      58.5%; Score 62; DB 4; Length 726;
Best Local Similarity 84.6%; Pred. No. 0.54;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DGKLYYDANSQD 13
Db 262 DGSRLYYDKNSGD 274

RESULT 9
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07

```

```

; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: fastseq for Windows Version 3.0
; SEQ ID NO 1
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: prt
; ORGANISM: Streptococcus mutans
US-09-740-274-2

```

Query Match 58.5%; Score 62; DB 3; Length 1475;  
Best Local Similarity 57.9%; Pred. NO. 1.2;  
Matches 11; Conservative 3; Mismatches 5; Indels

Qy 2 GKLRYYDANS GDQAFN KSV 20  
| : : | | | | | | : |  
pb 1264 GRISYYDGN SGDOIR NREV 1282

```

RESULT 10
US-10-383-930-34
; Sequence 34, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-34

```

Query Match	58.5%	Score 62;	DB 4;	Length 1475;
Best Local Similarity	57.9%	Pred. No. 1.2;		
Matches 11: Conservative		3: Mismatches	5: Indels	0: Gaps

Qy 2 GKLRYYDANSGDQAFNKSV 20  
|::||| ||| |:  
Db 1264 GRISYYDGNSGDQIRNRFV 1282

RESULT 11  
US-10-797-821-34  
; Sequence 34, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIORITY APPLICATION NUMBER: 10/383,930  
; PRIORITY FILING DATE: 2003-03-07  
; PRIORITY APPLICATION NUMBER: 60/363,209  
; PRIORITY FILING DATE: 2002-03-07  
; PRIORITY APPLICATION NUMBER: 60/402,483  
; PRIORITY FILING DATE: 2002-08-08  
; PRIORITY APPLICATION NUMBER: 09/290,049  
; PRIORITY FILING DATE: 1999-04-12  
; PRIORITY APPLICATION NUMBER: 60/081,550  
; PRIORITY FILING DATE: 1998-04-13  
; PRIORITY APPLICATION NUMBER: 60/115,142  
; PRIORITY FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45

```
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
UIS-10-797-821-34
```

Query Match 58.5%; Score 62; DB 5; Length 1475;  
Best Local Similarity 57.9%; Pred. No. 1.2;  
Matches 11; Conservative 3; Mismatches 5; Indels

QY 2 GKLRYYDANSGDQAFNKSV 20  
| : | | | | | : |  
DQ 1264 GRISYYDGNSGDOIRNFV 1282

```

RESULT 12
US-10-484-218-18
; Sequence 18, Application US/10484218
; Publication No. US20050059633A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERBITDINA HENDRIKA
; TITLE OF INVENTION: GLUCANS AND GLUCANUCRASES DERIVED FROM
; FILE OF INVENTION: LACTIC ACID BACTERIA
; FILE REFERENCE: 2001-1316
; CURRENT APPLICATION NUMBER: US/10/484,218
; CURRENT FILING DATE: 2004-01-20
; PRIOR APPLICATION NUMBER: PCT/NL02/00495
; PRIOR FILING DATE: 2002-07-22
; PRIOR APPLICATION NUMBER: EP 01202752.0
; PRIOR FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: EP 01202841.1
; PRIOR FILING DATE: 2001-07-25
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 18
; LENGTH: 1497
; TYPE: PRT
; ORGANISM: Lactobacillus sp.
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (284)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1345)
; OTHER INFORMATION: Variable amino acid
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (1453)
; OTHER INFORMATION: Variable amino acid
; US-10-484-218-18

```

Query Match 53.8%; Score 57; DB 5; Length 1497;  
Best Local Similarity 52.9%; Pred. No. 7.2;  
Matches 9: Conservative 3; Mismatches 5: Indels

Qy 1 DGKLRYYDANSQDQAFN 17  
: | | | : | | | : | |  
Db 1264 NGNLRVYFDGNTGDMVIN 1280

RESULT 13  
US-09-740-274-6  
Sequence 6, Application US/09740274  
Patent No. US20020031826A1  
GENERAL INFORMATION:  
APPLICANT: Nichols, Scott E.  
TITLE OF INVENTION: Glucan-containing Compositions and Paper  
FILE REFERENCE: 0357CRD  
CURRENT APPLICATION NUMBER: US/09/740,274  
CURRENT FILING DATE: 2000-12-19  
PRIORITY APPLICATION NUMBER: 09/310,361



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:15:51 ; Search time 0.866935 Seconds  
(without alignments)  
302.734 Million cell updates/sec

Title: US-10-797-821-27

Perfect score: 106

Sequence: 1 DGKLYVDANSQAFNKS 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62	58.5	726	7	US-11-052-554A-217
2	44	41.5	411	7	US-11-024-959-330
3	41	38.7	22	7	US-11-054-515-2169
4	41	38.7	258	7	US-11-054-515-862
5	41	38.7	358	6	US-10-689-742-170
6	41	38.7	398	6	US-10-873-328-61
7	40.5	38.2	1342	7	US-11-115-639-2
8	40.5	38.2	1342	7	US-11-115-639-3
9	40.5	38.2	1342	7	US-11-115-639-4
10	40.5	38.2	1342	7	US-11-115-639-5
11	40.5	38.2	1342	7	US-11-115-639-6
12	40	37.7	240	7	US-11-129-143-179
13	40	37.7	332	6	US-10-873-528-190
14	40	37.7	515	6	US-10-467-657-1968
15	40	37.7	719	6	US-10-467-657-762
16	39.5	37.3	608	7	US-11-024-959-481
17	39	36.8	243	7	US-11-054-515-2009
18	39	36.8	243	7	US-11-054-515-2012
19	39	36.8	350	6	US-10-878-556A-66
20	39	36.8	635	6	US-10-821-234-1573
21	39	36.8	2890	7	US-11-115-639-31
22	39	36.8	2890	7	US-11-115-639-32
23	39	36.8	2890	7	US-11-115-639-33
24	38.5	36.3	607	7	US-11-024-959-381
25	38	35.8	22	7	US-11-054-515-2165

ALIGNMENTS

RESULT 1

US-11-052-554A-217

; Sequence 217, Application US/11052554A

; Publication No. US20050288866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE

; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL

; FILE REFERENCE: 30853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; CURRENT FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 217

; LENGTH: 726

; TYPE: PRT

; ORGANISM: Streptococcus mutans UA159

US-11-052-554A-217

Query Match 58.5%; Score 62; DB 7; Length 726;

Best Local Similarity 84.6%; Pred. No. 0.016;

Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DGKLYVDANS 13

Db 262 DGSLRYDKNSGD 274

RESULT 2

US-11-024-959-330

; Sequence 330, Application US/11024959

; Publication No. US20060010516A1

; GENERAL INFORMATION:

; APPLICANT: FORSTER, RICHARD L.

; APPLICANT: CONNETT, MARIE B.

; APPLICANT: EMERSON, SARAH JANE

; APPLICANT: GRIGOR, MURRAY ROBERT

; APPLICANT: HIGGINS, COLLEEN M.

; APPLICANT: LUND, STEVEN TROY

; APPLICANT: MAGUSIN, ANDREAS

; APPLICANT: KODRZYCKI, BOB

; TITLE OF INVENTION: CELL CYCLE GENES AND RELATED METHODS

; FILE REFERENCE: 044463-0360

; CURRENT APPLICATION NUMBER: US/11/024,959



RESULT 10  
US-11-115-639-5  
; Sequence 5, Application US/11115639  
; Publication No. US20050282242A1  
GENERAL INFORMATION.

; APPLICANT: Rothstein, David  
; APPLICANT: Murphy, Christopher  
; APPLICANT: MacNeil, Ian  
; TITLE OF INVENTION: SCREENING ASSAYS FOR ANTIMICROBIAL  
; TITLE OF INVENTION: AGENTS  
; FILE REFERENCE: 50150/075003  
; CURRENT FILING DATE: 2005-04-27  
; PRIOR APPLICATION NUMBER: 60/566,858  
; PRIOR FILING DATE: 2004-04-30  
; PRIOR APPLICATION NUMBER: 60/565,679  
; PRIOR FILING DATE: 2004-04-27  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 1342  
; TYPE: PRT  
; ORGANISM: E. coli  
US-11-115-639-5

Query Match 38.2%; Score 40.5; DB 7; Length 1342;  
Best Local Similarity 42.1%; Pred. No. 90;  
Matches 8; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

QY 2 GKLYRYDANSQDQAFNKS 20  
Db 1208 GQIRLYDGRGEG-FERP 1225  
::|||::|||::|||::|||

RESULT 11  
US-11-115-639-6  
; Sequence 6, Application US/11115639  
; Publication No. US20050282242A1  
; GENERAL INFORMATION:  
; APPLICANT: Rothstein, David  
; APPLICANT: Murphy, Christopher  
; APPLICANT: MacNeil, Ian  
; TITLE OF INVENTION: SCREENING ASSAYS FOR ANTIMICROBIAL  
; TITLE OF INVENTION: AGENTS  
; FILE REFERENCE: 50150/075003  
; CURRENT FILING DATE: 2005-04-27  
; PRIOR APPLICATION NUMBER: 60/566,858  
; PRIOR FILING DATE: 2004-04-30  
; PRIOR APPLICATION NUMBER: 60/565,679  
; PRIOR FILING DATE: 2004-04-27  
; NUMBER OF SEQ ID NOS: 55  
; SOFTWARE: Fast-SEQ for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 1342  
; TYPE: PRT  
; ORGANISM: E. coli  
US-11-115-639-6

Query Match 38.2%; Score 40.5; DB 7; Length 1342;  
Best Local Similarity 42.1%; Pred. No. 90;  
Matches 8; Conservative 5; Mismatches 5; Indels 1; Gaps 1;

QY 2 GKLYRYDANSQDQAFNKS 20  
Db 1208 GQIRLYDGRGEG-FERP 1225  
::|||::|||::|||::|||

RESULT 12  
US-11-129-143-179  
; Sequence 179, Application US/11129143  
; Publication No. US20050266518A1  
; GENERAL INFORMATION:  
; APPLICANT: BERRY, Alan  
; APPLICANT: BRETZEL, Werner  
; APPLICANT: HUMBELIN, Markus  
; APPLICANT: LOPEZ-ULIBARRI, Rual  
; APPLICANT: MAYER, Anne F.

; APPLICANT: YELISEEV, Alexei A.  
; TITLE OF INVENTION: IMPROVED ISOPRENOID PRODUCTION  
; FILE REFERENCE: C38435/121966  
; CURRENT APPLICATION NUMBER: US/11/129,143  
; CURRENT FILING DATE: 2005-05-13  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 179  
; LENGTH: 240  
; TYPE: PRT  
; ORGANISM: Paracoccus sp. R114  
US-11-129-143-179

Query Match 37.7%; Score 40; DB 7; Length 240;  
Best Local Similarity 42.1%; Pred. No. 19;  
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 2 GKLYRYDANSQDQAFNKS 20  
Db 143 GQANYAAKAGDILGFTKSL 161  
::|||::|||::|||::|||

RESULT 13  
US-10-873-528-190  
; Sequence 190, Application US/10873528  
; Publication No. US20050276814A1  
; GENERAL INFORMATION:  
; APPLICANT: Microbial Technics Limited  
; APPLICANT: Gilbert, Christophe FG  
; APPLICANT: Hansbro, Philip M  
; TITLE OF INVENTION: Proteins  
; FILE REFERENCE: PWC/P21129W0  
; CURRENT APPLICATION NUMBER: US/10/873,528  
; CURRENT FILING DATE: 2004-06-23  
; PRIOR APPLICATION NUMBER: US/09/769,787  
; PRIOR FILING DATE: 2001-01-26  
; PRIOR APPLICATION NUMBER: GB 9816337.1  
; PRIOR FILING DATE: 1998-03-27  
; PRIOR APPLICATION NUMBER: US 60/125164  
; PRIOR FILING DATE: 1999-03-19  
; NUMBER OF SEQ ID NOS: 388  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 190  
; LENGTH: 332  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-10-873-528-190

Query Match 37.7%; Score 40; DB 6; Length 332;  
Best Local Similarity 53.8%; Pred. No. 26;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 DGKLYRYDANSQD 13  
Db 58 DGKYYVFDLSGE 70  
::|||::|||::|||::|||

RESULT 14  
US-10-467-657-1968  
; Sequence 1968, Application US/10467657  
; Publication No. US20050260581A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SPA  
; APPLICANT: FONTANA Maria Rita  
; APPLICANT: PIZZA Mariagrazia  
; APPLICANT: MASIGNANI Vega  
; APPLICANT: MONACI Elisabetta  
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/467,657  
; CURRENT FILING DATE: 2003-08-11  
; PRIOR APPLICATION NUMBER: GB-0103424.8  
; PRIOR FILING DATE: 2001-02-12



; NUMBER OF SEQ ID NOS: 9218  
; SOFTWARE: SeqWin99, version 1.04  
; SEQ ID NO 1968  
; LENGTH: 515  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-1968

Query Match 37.7%; Score 40; DB 6; Length 515;  
Best Local Similarity 55.6%; Pred. No. 41;  
Matches 10; Conservative 1; Mismatches 5; Indels 2; Gaps 1;

Qy 1 DGKLYDYDANSQDAFNK 18  
||| ||| ||| |||  
Db 140 DGNL--YWAGSGDFVFNQ 155

RESULT 15

US-10-467-657-762  
; Sequence 762, Application US/10467657  
; Publication No. US20050260581A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SPA  
; APPLICANT: FONTANA Maria Rita  
; APPLICANT: PIZZA Mariagrazia  
; APPLICANT: MASIGNANI Vega  
; APPLICANT: MONACI Elisabetta  
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/467,657  
; CURRENT FILING DATE: 2003-08-11  
; PRIOR APPLICATION NUMBER: GB-0103424.8  
; PRIOR FILING DATE: 2001-02-12  
; NUMBER OF SEQ ID NOS: 9218  
; SOFTWARE: SeqWin99, version 1.04  
; SEQ ID NO 762  
; LENGTH: 719  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-762

Query Match 37.7%; Score 40; DB 6; Length 719;  
Best Local Similarity 44.4%; Pred. No. 57;  
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Qy 1 DGKLYDYDANSQDAFNK 18  
||| ||| ||| |||  
Db 700 DGKLLDLDVDAQPGDK 717

Search completed: February 10, 2006, 23:26:02  
Job time : 1.86694 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:25:38 ; Search time 200 Seconds

(without alignments)  
30.757 Million cell updates/sec

Title: US-10-797-821-28

Perfect score: 85

Sequence: 1 QWNGSEKPYDDHL 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : A\_Geneseq\_21.\*

1: Geneseqp1980s.\*

2: Geneseqp1990s.\*

3: Geneseqp2000s.\*

4: Geneseqp2001s.\*

5: Geneseqp2002s.\*

6: Geneseqp2003as.\*

7: Geneseqp2003bs.\*

8: Geneseqp2004s.\*

9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	85	100.0	14	2 AAR43698	Aar43698 GT subseq
2	85	100.0	14	2 AAW34160	Aaw34160 GTF antig
3	85	100.0	14	2 AAY43339	Aay43339 GTF antig
4	85	100.0	14	7 ADD93648	Add93648 Streptoco
5	85	100.0	14	9 ADX37271	Adx37271 Streptoco
6	85	100.0	15	9 ADV68559	Adv68559 S. sobrin
7	85	100.0	15	9 ADV68558	Adv68558 S. downei
8	85	100.0	15	9 ADV68554	Adv68554 Glucosylt
9	85	100.0	16	2 AAW34165	Aaw34165 GTF antig
10	85	100.0	17	5 ABB98615	Abb98615 Dextrane-
11	85	100.0	1590	7 ADD93657	Add93657 Streptoco
12	85	100.0	1590	9 ADX37280	Adx37280 Streptoco
13	85	100.0	1592	2 AAR43295	Aar43295 Glucosylt
14	67	78.8	15	2 AAY43354	Aay43354 GTF antig
15	67	78.8	15	9 ADV68556	Adv68556 S. mutans
16	67	78.8	17	5 ABB98614	Abb98614 Dextrane-
17	67	78.8	126	2 AAR12630	Aar12630 GtFB.1/CT
18	67	78.8	1017	5 AAU79285	Aau79285 Streptoco
19	67	78.8	1375	5 AAU98028	Aau98028 S. mutans
20	67	78.8	1375	5 AAU79288	Aau79288 Streptoco
21	67	78.8	1375	7 ADD93655	Add93655 Streptoco
22	67	78.8	1375	9 ADX37278	Adx37278 Streptoco
23	67	78.8	1475	5 AAU98031	Aau98031 S. mutans
24	67	78.8	1475	5 AAU98036	Aau98036 S. mutans

25	67	78.8	1475	5 AAU98037	Aau98037 S. mutans
26	67	78.8	1475	5 AAU98040	Aau98040 S. mutans
27	67	78.8	1475	5 AAU98035	Aau98035 S. mutans
28	67	78.8	1475	5 AAU98033	Aau98033 S. mutans
29	67	78.8	1475	5 AAU98034	Aau98034 S. mutans
30	67	78.8	1475	5 AAU98030	Aau98030 S. mutans
31	67	78.8	1475	5 AAU98039	Aau98039 S. mutans
32	67	78.8	1475	5 AAU98027	Aau98027 S. mutans
33	67	78.8	1475	5 AAU98032	Aau98032 S. mutans
34	67	78.8	1475	5 AAU98038	Aau98038 S. mutans
35	67	78.8	1475	7 ADD93654	Add93654 Streptoco
36	67	78.8	1475	9 ADX37277	Adx37277 Streptoco
37	67	78.8	1475	5 AAU79284	Aau79284 Streptoco
38	64	75.3	15	9 ADV68555	Adv68555 S. mutans
39	51	60.0	17	5 ABB98619	Abb98619 Dextrane-
40	51	60.0	1477	9 ADY72733	Ady72733 Mutant de
41	51	60.0	1477	9 ADY72732	Ady72732 Mutant de
42	51	60.0	1477	9 ADY72696	Ady72696 Mutant de
43	49	57.5	102	4 AAM86301	Aam86301 Human imm
44	48	56.5	105	6 ABUS0118	Abu50118 Protein e
45	47.5	55.9	357	7 ADF04672	Adf04672 Bacterial
46	47	55.3	105	6 ABU40591	Abu40591 Protein e
47	47	55.3	106	7 ADF08022	Adf08022 Bacterial
48	47	55.3	1149	6 ABR63236	Abr63236 Glucanau
49	46	54.1	138	4 AAM80111	Aam80111 Human pro
50	46	54.1	504	5 ABP38135	Abp38135 Staphyloc
51	46	54.1	504	8 ADS07661	Ads07661 Staphyloc
52	45	52.9	105	6 ABU49626	Abu49626 Protein e
53	44	51.8	480	6 ADY09855	Ady09855 Plant ful
54	43	50.6	17	5 ABB98618	Abb98618 Dextrane-
55	43	50.6	347	8 ADQ16288	Adq16288 Pea proba
56	42.5	50.0	268	4 ABB67594	Abb67594 Drosophil
57	42	49.4	17	5 ABB98617	Abb98617 Dextrane-
58	42	49.4	70	4 AAU45650	Aau45650 Propionib
59	42	49.4	70	6 ABM42169	Abm42169 Propionib
60	42	49.4	105	3 AAB15396	Aab15396 E. coli m
61	42	49.4	105	6 ABU48079	Abu48079 Protein e
62	42	49.4	105	6 ABU31935	Abu31935 Protein e
63	42	49.4	105	6 ABU28388	Abu28388 Protein e
64	42	49.4	105	6 ABU45384	Abu45384 Protein e
65	42	49.4	105	6 ABU15170	Abu15170 Protein e
66	42	49.4	187	7 ABO65685	Abob65685 Klebsiell
67	42	49.4	188	4 ABG04553	Abg04553 Novel hum
68	42	49.4	271	7 ABO81987	Aob81987 Pseudomon
69	42	49.4	299	4 ABG30019	Abg30019 Novel hum
70	42	49.4	424	3 AAB28785	Aab28785 Streptoco
71	42	49.4	424	4 AAU37873	Aau37873 Streptoco
72	42	49.4	424	4 AAB60871	Aab60871 Mevalonat
73	42	49.4	424	4 AAM01110	Aam01110 CFE 113 p
74	42	49.4	424	6 ABU02203	Abu02203 S. pneumo
75	42	49.4	424	6 ABU46206	Abu46206 Protein e
76	42	49.4	424	8 ADK46604	Adk46604 Streptoco
77	42	49.4	424	8 ADM98850	Adm98850 HMG-CoA r
78	42	49.4	424	8 ADM98751	Adm98751 HMG-CoA r
79	42	49.4	424	8 ADU80769	Adu80769 Streptoco
80	42	49.4	426	8 ADR95055	Adr95055 Novel S.
81	42	49.4	426	9 AEA58925	Aea58925 Streptoco
82	42	49.4	1429	8 ADN23166	Adn23166 Bacterial
83	42	49.4	1527	5 AAU80055	Aau80055 Leuconost
84	42	49.4	1527	7 ADC54807	Adc54807 Leuconost
85	41.5	48.8	54	4 AAU41100	Aau41100 Propionib
86	41.5	48.8	54	6 ABM37619	Abm37619 Propionib
87	41	48.2	80	4 AAM87305	Aam87305 Human imm
88	41	48.2	304	8 ADM87713	Adm87713 Human EST
89	41	48.2	363	6 AAO16338	Aao16338 Ochrobact
90	41	48.2	367	9 ADY65206	Ady65206 S. manson
91	41	48.2	471	6 ABU40741	Abu40741 Protein e
92	41	48.2	473	7 ADF04195	Adf04195 Bacterial
93	41	48.2	488	8 ADN21027	Adn21027 Bacterial
94	41	48.2	1202	7 ADE15630	Ade15630 Human str
95	41	48.2	1202	8 ADM87259	Adm87259 Human pro
96	40	47.1	204	3 AAG33294	Aag33294 Zea may
97	40	47.1	238	8 ADY12632	Ady12632 Plant ful

98	40	47.1	242	4	ABG05940	Abg05940 Novel hum	171	39	45.9	1639	2	AAW54145	Aaw54145 P. falcip
99	40	47.1	252	7	ADH85844	Adh85844 Enterococ	172	39	45.9	1639	5	AEE29345	Aee29345 Plasmodiu
100	40	47.1	269	3	RAG33293	Rag33293 Zea mays	173	39	45.9	1654	1	AAE50777	Aae50777 Sequence
101	40	47.1	286	7	ABM73628	Abm73628 DNA clone	174	39	45.9	1771	7	ADM26928	Adm26928 Hyperther
102	40	47.1	304	8	ADX66011	Adx66011 Plant ful	175	39	45.9	2835	5	ABR98574	Abrr98574 Dextran s
103	40	47.1	319	4	AAG82833	Agag82833 S. epider	176	39	45.9	2835	6	ABR55594	Abrr55594 Amino aci
104	40	47.1	326	3	RAG35957	Rag35957 Zea mays	177	38.5	45.3	357	6	AUA45695	Aua45695 Propionib
105	40	47.1	316	8	ADG90128	Adg90128 Plant ful	178	38.5	45.3	357	6	ABM42214	Abm42214 Propionib
106	40	47.1	423	8	ADY13589	Ady13589 Plant ful	179	38.5	45.3	452	9	AEB90523	Aeb90523 Thielavia
107	40	47.1	423	8	ADX96394	Adx96394 Plant ful	180	38.5	45.3	603	9	ABE27323	Abe27323 Pinus rad
108	40	47.1	436	6	ABU42678	Abu42678 Protein e	181	38	44.7	69	8	ABO54638	Abos4638 Human gen
109	40	47.1	440	3	RAG35956	Rag35956 Zea mays	182	38	44.7	80	7	ADC00984	Adc00984 Enterohae
110	40	47.1	443	5	ABP38338	Abp38338 Staphyloc	183	38	44.7	81	8	ADM66580	Adm66580 PRRSV Gp5
111	40	47.1	443	3	ADS04863	Ads04863 Staphyloc	184	38	44.7	82	8	ADN33136	Adn33136 Human tra
112	40	47.1	463	3	RAG35955	Rag35955 Zea mays	185	38	44.7	108	7	ADF13855	Adf13855 Human end
113	40	47.1	466	8	ADT60853	Adt60853 Plant pol	186	38	44.7	108	7	ADF13856	Adf13856 Human end
114	40	47.1	474	8	ADY09814	Ady09814 Plant ful	187	38	44.7	124	7	ADH86625	Adh86625 Enterococ
115	40	47.1	513	4	ABB70632	Abb70632 Drosophil	188	38	44.7	144	8	ADL05289	Adl05289 M. catarr
116	40	47.1	554	6	ABU40161	Abu40161 Protein e	189	38	44.7	170	8	ADX74554	Adx74554 Plant ful
117	40	47.1	554	6	ABU40144	Abu40144 Protein e	190	38	44.7	196	4	AAU03573	Aau03573 P. gingiv
118	40	47.1	554	9	ADM11173	Adm11173 Helicobac	191	38	44.7	198	6	ABU23574	Abu23574 Protein e
119	40	47.1	606	8	ADX66338	Adx66338 Plant ful	192	38	44.7	207	8	ADX76400	Adx76400 Plant ful
120	40	47.1	1001	6	ABR52956	Abrr52956 Protein s	193	38	44.7	212	3	AGS7490	Agss7490 Arabidops
121	40	47.1	1001	7	ADK62626	Adk62626 Disease t	194	38	44.7	213	3	AGG60522	Agg60522 Arabidops
122	40	47.1	1001	7	ADS43914	Ads43914 Bacterial	195	38	44.7	218	3	AGS7489	Agss7489 Arabidops
123	40	47.1	1006	8	ADN21545	Adn21545 Bacterial	196	38	44.7	219	3	AGG60521	Agg60521 Arabidops
124	40	47.1	2257	4	ABB61782	Abb61782 Drosophil	197	38	44.7	222	3	AGG11357	Agg11357 Arabidops
125	39.5	46.5	58	4	ABR39754	Abbr39754 Peptide #	198	38	44.7	222	8	ADS24180	Ads24180 Bacterial
126	39.5	46.5	58	4	AAAM3336	Aaam3336 Peptide #	199	38	44.7	254	8	ADY09453	Ady09453 Human pro
127	39.5	46.5	58	4	AAAM73119	Aaam73119 Human bon	200	38	44.7	258	8	ADY07856	Ady07856 Plant ful
128	39.5	46.5	58	4	AAAG60474	Aaag60474 Human bra	201	38	44.7	269	7	ADT07193	Adt07193 Bacterial
129	39.5	46.5	58	4	ABGS4835	Abgs4835 Human liv	202	38	44.7	274	8	ADT59354	Adt59354 Plant pol
130	39.5	46.5	58	5	ABG42965	Abg42965 Human pep	203	38	44.7	288	3	AGS7488	Agss7488 Arabidops
131	39.5	46.5	194	8	ADX72617	Adx72617 Plant ful	204	38	44.7	288	3	AGG11356	Agg11356 Arabidops
132	39.5	46.5	414	8	ADI00603	Adi00603 Human up-	205	38	44.7	289	3	AGG60520	Agg60520 Arabidops
133	39.5	46.5	465	4	ABG20190	Abg20190 Novel hum	206	38	44.7	349	3	AAO78823	Aao78823 Amino aci
134	39.5	46.5	1693	4	ABBS8927	Abbs8927 Drosophil	207	38	44.7	353	3	AAO78821	Aao78821 Amino aci
135	39	45.9	17	5	ABBS98621	Abbs98621 Dextrane-	208	38	44.7	358	6	ABM67478	Abm67478 Photorhab
136	39	45.9	17	5	ABBS98575	Abbs98575 Dextrane	209	38	44.7	358	8	ADU07215	Adu07215 Paspalum
137	39	45.9	17	6	ABRS5595	Abrr5595 Peptide d	210	38	44.7	380	5	AAU09398	Aau09398 Soluble P
138	39	45.9	51	4	AAU45414	Aau45414 Propionib	211	38	44.7	380	7	ABW00418	Abw00418 Porphyrom
139	39	45.9	51	6	ABM41933	Abm41933 Propionib	212	38	44.7	385	2	AAV34491	Aav34491 Porphyrom
140	39	45.9	89	4	AAO12302	Aao12302 Human pol	213	38	44.7	385	7	ABW00417	Abw00417 Porphyrom
141	39	45.9	157	6	ABU20373	Abu20373 Protein e	214	38	44.7	387	2	AAV34366	Aav34366 Porphyrom
142	39	45.9	203	7	ADC15603	Adc15603 Murine de	215	38	44.7	391	8	ABM82195	Abm82195 Tumour-as
143	39	45.9	218	4	AAU51621	Aau51621 Propionib	216	38	44.7	403	8	ADX96536	Adx96536 Plant ful
144	39	45.9	218	6	ABM48140	Abm48140 Propionib	217	38	44.7	408	8	ADX92941	Adx92941 Plant ful
145	39	45.9	222	4	AAU67904	Aau67904 Propionib	218	38	44.7	408	8	ADX93535	Adx93535 Plant ful
146	39	45.9	222	6	ABM64423	Abm64423 Propionib	219	38	44.7	413	5	ABP73229	Abp73229 Candida a
147	39	45.9	237	7	ADC15599	Adc15599 Murine de	220	38	44.7	414	7	ABM89508	Abm89508 Rice abio
148	39	45.9	267	8	ADR51415	Adr51415 Anti-biof	221	38	44.7	431	5	AAU74641	Aau74641 Oestrogen
149	39	45.9	309	4	AGG81631	Agg81631 S. epider	222	38	44.7	443	7	ABO72510	Abos72510 Pseudomon
150	39	45.9	313	5	ABP39526	Abp39526 Staphyloc	223	38	44.7	444	7	ABM90244	Abm90244 Rice abio
151	39	45.9	313	8	ADS05459	Ads05459 Staphyloc	224	38	44.7	445	3	ABM16689	Abm16689 Bacteriop
152	39	45.9	339	5	ABP30411	Abp30411 Streptoco	225	38	44.7	509	6	ADA54608	Ada54608 Human pro
153	39	45.9	339	5	ABR53330	Abrr53330 Protein s	226	38	44.7	532	7	ADA08081	Ada08081 Human PK2
154	39	45.9	339	7	ADK63452	Adk63452 Disease t	227	38	44.7	532	8	ADJ32261	Adj32261 Human PK2
155	39	45.9	347	5	ABP29900	Abp29900 Streptoco	228	38	44.7	599	6	ABP58340	Abp58340 Human cel
156	39	45.9	347	5	ABP28641	Abp28641 Streptoco	229	38	44.7	599	7	ADM05328	Adm05328 Human pro
157	39	45.9	347	8	ADV82367	Adv82367 Streptoco	230	38	44.7	640	8	ADN22573	Adn22573 Bacterial
158	39	45.9	360	5	AAE29347	Aae29347 Plasmodiu	231	38	44.7	669	4	ABG26242	Abg26242 Novel hum
159	39	45.9	366	7	ADH52742	Adh52742 Porphyrom	232	38	44.7	750	5	ADJ33826	Adj33826 Protein o
160	39	45.9	422	4	ABG07995	Abg07995 Novel hum	233	38	44.7	863	3	AAB35756	Aab35756 ETAF env
161	39	45.9	440	8	ADP80748	Adp80748 T pacific	234	38	44.7	865	4	AGG67876	Agg67876 Equine in
162	39	45.9	548	7	ADJ80137	Adj80137 Novel hum	235	38	44.7	1003	2	AAW35003	Aaw35003 Thermotog
163	39	45.9	561	8	ADS23379	Ads23379 Bacterial	236	38	44.7	1010	8	ADN20415	Adn20415 Bacterial
164	39	45.9	566	8	ADS25221	Ads25221 Bacterial	237	38	44.7	1018	6	ABU35384	Abu35384 Proteinh e
165	39	45.9	622	8	ADX89116	Adx89116 Plant ful	238	38	44.7	1095	2	AAW17888	Aaw17888 Photorhab
166	39	45.9	648	8	ADS22420	Ads22420 Bacterial	239	38	44.7	1095	2	AAW56547	Aaw56547 Toxin Tca
167	39	45.9	671	8	ADN25746	Adn25746 Bacterial	240	38	44.7	1095	6	ABM67946	Abm67946 Photorhab
168	39	45.9	711	8	ADR86405	Adr86405 Aspergill	241	38	44.7	1301	2	AAR28304	Aar28304 Acid alph
169	39	45.9	911	4	AAU56103	Aau56103 Propionib	242	38	44.7	1312	4	ABW60755	Abw60755 Drosophil
170	39	45.9	911	6	ABM52622	Abm52622 Propionib	243	37.5	44.1	168	2	AAW24127	Aaw24127 Pyrococcu

244	37.5	44.1	238	8	ADY13014	Ady13014 Plant ful	317	37	43.5	500	2	AAy04867	Mycobacte
245	37.5	44.1	312	3	AAg59983	Arabidops	318	37	43.5	500	8	ADU16770	M. tuberc
246	37.5	44.1	332	6	ABU97164	Enzyme po	319	37	43.5	502	7	ABO74755	Pseudomon
247	37.5	44.1	370	3	AAg59982	Arabidops	320	37	43.5	505	8	ADG73740	Aspergill
248	37.5	44.1	381	3	AAg59981	Arabidops	321	37	43.5	511	7	ADC12752	Human GPC
249	37.5	44.1	412	2	AAW94836	Hyperther	322	37	43.5	515	9	AEa81671	Mouse ika
250	37.5	44.1	488	9	ABE27313	Pinus rad	323	37	43.5	522	8	ADx76208	Plant ful
251	37.5	44.1	522	2	AAW24122	Pyrococcu	324	37	43.5	527	8	AAy04868	Mycobacte
252	37.5	44.1	522	2	AAW94838	Hyperther	325	37	43.5	527	8	ADU16772	M. tuberc
253	37.5	44.1	654	2	AAW24129	Pyrococcu	326	37	43.5	539	8	ADs24692	Bacterial
254	37.5	44.1	654	2	AAW94841	Hyperther	327	37	43.5	547	5	AAU10669	Mouse LiC
255	37.5	44.1	659	2	AAW24123	Protease.	328	37	43.5	554	6	ABU38793	Protein e
256	37.5	44.1	882	4	ABE58526	Drosophil	329	37	43.5	554	6	ABU41346	Protein e
257	37.5	44.1	936	4	ABG20576	Novel hum	330	37	43.5	558	4	ABG12118	Novel hum
258	37	43.5	36	5	ABP29510	Streptoco	331	37	43.5	565	7	ABO78276	Pseudomon
259	37	43.5	70	4	AAW21267	Peptide #	332	37	43.5	611	4	AAAB96570	Putative
260	37	43.5	70	4	ABB43600	Peptide #	333	37	43.5	624	4	ABG07605	Novel hum
261	37	43.5	70	4	AAW37494	Peptide #	334	37	43.5	625	6	ABU41656	Protein e
262	37	43.5	70	4	ABB25554	Protein #	335	37	43.5	631	4	ABG11126	Novel hum
263	37	43.5	70	4	AAW77344	Human bon	336	37	43.5	632	7	ABM89322	Rice abio
264	37	43.5	70	4	AAW64539	Human bra	337	37	43.5	635	5	AAU10668	Mouse LiC
265	37	43.5	70	4	ABG58976	Human liv	338	37	43.5	636	7	ADJ70197	Human hea
266	37	43.5	70	5	ABG46359	Human pep	339	37	43.5	650	8	ADS29636	Bacterial
267	37	43.5	91	3	ABA41130	Human ORF	340	37	43.5	684	9	AEa25767	Executor
268	37	43.5	92	9	AEA25762	Executor	341	37	43.5	695	7	ADL02110	Guinea pi
269	37	43.5	101	8	ABO55179	Human gen	342	37	43.5	714	3	AAG42605	Arabidops
270	37	43.5	104	4	ABG26849	Novel hum	343	37	43.5	743	8	ABM83738	Human dia
271	37	43.5	111	4	AAO00158	Human pol	344	37	43.5	746	4	ABG07028	Novel hum
272	37	43.5	117	9	AEA25760	Executor	345	37	43.5	749	4	AAG62727	Novel aci
273	37	43.5	126	4	AAW16771	Peptide #	346	37	43.5	749	7	ADD08936	Human sem
274	37	43.5	126	4	ABB35757	Peptide #	347	37	43.5	749	7	ADL82793	Human sem
275	37	43.5	126	4	AAW29258	Peptide #	348	37	43.5	749	8	ABM80457	Tumour-as
276	37	43.5	126	4	ABB30591	Peptide #	349	37	43.5	749	9	ADZ00428	Senaphori
277	37	43.5	126	4	ABB21181	Protein #	350	6	ABU38870	Protein e			
278	37	43.5	126	4	AAW68947	Human bon	351	37	43.5	753	8	ADU07705	Amino aci
279	37	43.5	126	4	AAW56565	Human bra	352	37	43.5	757	3	AAG42604	Arabidops
280	37	43.5	126	4	ABG50610	Human liv	353	37	43.5	757	5	ABB1980	Herbicida
281	37	43.5	126	4	AAW04488	Peptide #	354	37	43.5	781	7	ABO74607	Pseudomon
282	37	43.5	126	5	ABG38531	Human pep	355	37	43.5	801	5	ABP65382	Bifidobac
283	37	43.5	147	4	ABG01127	Novel hum	356	37	43.5	806	5	ABB93653	Herbicida
284	37	43.5	184	6	ABU44515	Protein e	357	37	43.5	820	8	ABM83737	Human dia
285	37	43.5	209	5	AAU10658	Mouse LiC	358	37	43.5	849	8	AD116358	Human pro
286	37	43.5	220	7	ABO72903	Pseudomon	359	37	43.5	858	8	ABM83736	Human dia
287	37	43.5	227	7	ABM86426	Rice abio	360	37	43.5	859	8	ABM83735	Human dia
288	37	43.5	228	9	AEA25758	Executor	361	37	43.5	860	8	ABM84932	Human dia
289	37	43.5	233	8	ADY13287	Plant ful	362	37	43.5	873	4	AAW79221	Human pro
290	37	43.5	238	4	AAU55019	Propionib	363	37	43.5	892	8	ABM83734	Human dia
291	37	43.5	238	6	ABM51538	Propionib	364	37	43.5	902	8	ABM83733	Human dia
292	37	43.5	251	8	ADX87348	Plant ful	365	37	43.5	910	8	ABM83731	Human dia
293	37	43.5	295	3	AAG23507	Arabidops	366	37	43.5	910	8	ABM83732	Human dia
294	37	43.5	296	3	AAW94956	M. xanthu	367	37	43.5	921	8	ABM83729	Human dia
295	37	43.5	301	9	ABM94956	Arabidops	368	37	43.5	922	8	ABM83730	Human dia
296	37	43.5	302	4	AAE00230	Protein e	369	37	43.5	922	8	ABM83728	Human dia
297	37	43.5	302	7	ADB91919	Acyltrans	370	37	43.5	923	4	AAW80205	Human pro
298	37	43.5	309	5	AAU10657	Mouse LiC	371	37	43.5	941	4	ABG21485	Novel hum
299	37	43.5	309	9	ADY59313	Mouse SMC	372	37	43.5	947	7	ABO72589	Pseudomon
300	37	43.5	348	6	ABM73426	Staphyloc	373	37	43.5	990	8	ADS34943	Human aut
301	37	43.5	351	5	ABB92633	Herbicida	374	37	43.5	990	8	ADS34942	Human aut
302	37	43.5	360	2	AAW76477	Recombina	375	37	43.5	1034	7	ABO72223	Pseudomon
303	37	43.5	367	7	ADE58275	Rat Prote	376	37	43.5	1041	7	ABO80520	Pseudomon
304	37	43.5	386	4	ABG18957	Novel hum	377	37	43.5	1065	2	AAW35852	Human CD1
305	37	43.5	397	5	AAU10656	Mouse LiC	378	37	43.5	1065	6	ABU05074	Human exp
306	37	43.5	425	8	ADT60859	Plant pol	379	37	43.5	1260	5	AAU10649	Mouse LiC
307	37	43.5	436	3	AAG23506	Arabidops	380	37	43.5	1260	9	AEA55057	Mouse neu
308	37	43.5	437	3	AAW48850	Arabidops	381	37	43.5	1260	9	AEA55087	Mouse Li
309	37	43.5	441	7	ADF04733	Bacterial	382	37	43.5	1419	4	ABE59208	Drosophil
310	37	43.5	447	5	AAU10670	Mouse LiC	383	37	43.5	1465	8	ADU07767	Amino aci
311	37	43.5	455	5	ABB91703	Herbicida	384	37	43.5	1563	7	ADK40927	Novel hum
312	37	43.5	469	8	ADS24848	Bacterial	385	37	43.5	1563	8	ADR15648	Kinase 47
313	37	43.5	478	7	ADE28110	Human NTR	386	37	43.5	1688	6	ABR58344	XM_047995
314	37	43.5	479	4	ABE59767	Drosophil	387	37	43.5	1737	8	ADK52107	Human ato
315	37	43.5	479	6	ABJ25983	Aspergill	388	37	43.5	1915	7	ADJ71130	Human hea
316	37	43.5	479	6	ABJ25383	Aspergill	389	37	43.5	1967	6	ABP81700	Human G P

390	37	43.5	2144	8	ADQ19484	Adq19484 Human sof	463	36	42.4	178	5	AAU99737	Aau99737 Pyrodicti
391	37	43.5	2333	8	ADH71274	Adh71274 Human pro	464	36	42.4	178	7	ADB46048	Adb46048 CanC fusi
392	37	43.5	2376	8	ADU07769	Adu07769 Amino aci	465	36	42.4	181	4	ABG18531	Abg18531 Novel hum
393	37	43.5	2551	6	ABR58317	AbR58317 BC02005A	466	36	42.4	181	5	ABP42890	Abp42890 Human ova
394	37	43.5	2633	6	ABR58318	AbR58318 BC02005B	467	36	42.4	183	4	ABG06676	Abg06676 Novel hum
395	37	43.5	2662	8	ADH71218	Adh71218 Human pro	468	36	42.4	183	8	ADS43551	Ads43551 Bacterial
396	37	43.5	2724	4	AAU08681	Aau08681 Human FCT	469	36	42.4	192	4	AAW81690	Aaw81690 Human hae
397	37	43.5	2724	7	ADB32029	AdB32029 Human FCT	470	36	42.4	192	4	AAW81426	Aam81426 Human hae
398	37	43.5	2724	8	ADH71252	Adh71252 Human pro	471	36	42.4	193	8	ADX90439	Adx90439 Plant ful
399	37	43.5	2733	4	AAU08680	Aau08680 Human FCT	472	36	42.4	193	8	ABO60083	Abo60083 Human gen
400	37	43.5	2733	7	ADB32024	AdB32024 Human FCT	473	36	42.4	202	7	ADF59544	Adf59544 Human pol
401	37	43.5	2733	8	ADH71250	Adh71250 Human pro	474	36	42.4	219	4	AAW72164	Aay72164 Human RNA
402	37	43.5	2733	8	ADH71254	Adh71254 Human pro	475	36	42.4	243	2	AAW38542	Aaw38542 Streptoco
403	37	43.5	2733	8	ADH71240	Adh71240 Human pro	476	36	42.4	248	9	AEB91360	Aeb91360 Microbial
404	37	43.5	2733	8	ADH71246	Adh71246 Human pro	477	36	42.4	248	9	AEB91632	Aeb91632 Microbial
405	37	43.5	2733	8	ADH71258	Adh71258 Human pro	478	36	42.4	258	8	ADQ67791	Adq67791 Novel hum
406	37	43.5	2759	5	ADI16608	Adi16608 Human NOV	479	36	42.4	266	7	ADC32753	Adc32753 Human nov
407	37	43.5	2759	8	ADH71272	Adh71272 Human pro	480	36	42.4	269	3	AAW84676	Aay84676 Amino aci
408	37	43.5	2764	5	ADI16951	Adi16951 Murine NO	481	36	42.4	271	6	ABM68045	Abm68045 Photorhab
409	37	43.5	2764	8	ADJ76262	Adj76262 Marker ge	482	36	42.4	275	8	ADX75415	Adx75415 Plant ful
410	37	43.5	2765	5	ADI16952	Adi16952 Rat NOVX	483	36	42.4	283	5	ABB78313	Abb78313 Amino aci
411	37	43.5	2765	8	ADH71220	Adh71220 Human pro	484	36	42.4	285	5	ABG93284	Abg93284 C. albica
412	37	43.5	2769	8	ADN42262	Adn42262 Human nov	485	36	42.4	295	8	ADX92156	Adx92156 Plant ful
413	37	43.5	2802	5	ADI16953	Adi16953 Chicken N	486	36	42.4	303	4	AAU18525	Aau18525 Human cyt
414	37	43.5	3530	4	AAW25586	Aam25586 Human pro	487	36	42.4	303	4	AAU87537	Aau87537 Novel cen
415	37	43.5	4131	8	ADN22483	Adn22483 Bacterial	488	36	42.4	303	8	ADI54852	Adi54852 Novel hum
416	37	43.5	6304	6	ABU09236	Abu09236 Human neu	489	36	42.4	307	8	ADX74849	Adx74849 Plant ful
417	37	43.5	6306	8	ADU00113	Adu00113 Amino aci	490	36	42.4	311	4	AAW14618	Aam14618 Peptide #
418	37	43.5	6307	8	ADO29686	Ado29686 Human GPC	491	36	42.4	311	4	ABB33578	Abb33578 Peptide #
419	37	43.5	6307	9	ADX83195	Adx83195 Human TEG	492	36	42.4	311	4	AAW27039	Aam27039 Peptide #
420	36.5	42.9	150	6	ADB09695	Adb09695 Allotococ	493	36	42.4	311	4	ABB28398	Abb28398 Peptide #
421	36.5	42.9	179	5	ABW78686	Abw78686 C. elegan	494	36	42.4	311	4	ABB19035	Abb19035 Protein #
422	36.5	42.9	187	5	ABB78688	Abb78688 C. elegan	495	36	42.4	311	4	AAW66752	Aam66752 Human bon
423	36.5	42.9	188	5	ABBS3836	Abbs3836 Lactococc	496	36	42.4	311	4	AAW54353	Aam54353 Human bra
424	36.5	42.9	202	6	ADX09693	Adx09693 Allotococ	497	36	42.4	311	4	ABG48421	Abg48421 Human liv
425	36.5	42.9	228	8	ADX92795	Adx92795 Plant ful	498	36	42.4	311	4	AAW02343	Aam02343 Peptide #
426	36.5	42.9	266	7	ADFI18526	Adfi18526 Haloalkan	499	36	42.4	311	5	ABG36407	Abg36407 Human pep
427	36.5	42.9	276	7	ADFI18517	Adfi18517 Haloalkan	500	36	42.4	317	2	AAW85961	Aay85961 S. pneumo
428	36.5	42.9	293	5	ABG98040	Abg98040 Rhodococc	501	36	42.4	317	3	AAW13516	Aab13516 Streptoco
429	36.5	42.9	293	7	ADFI18513	Adfi18513 Rhodococc	502	36	42.4	317	4	AAU38076	Aau38076 Streptoco
430	36.5	42.9	296	8	ADW2533	Adw2533 Halo-alka	503	36	42.4	317	4	AAU37833	Aau37833 Streptoco
431	36.5	42.9	306	5	ABG98042	Abg98042 Mutant ha	504	36	42.4	317	4	ABG60873	Abg60873 Mevalonat
432	36.5	42.9	307	5	ABG98041	Abg98041 Mycobacte	505	36	42.4	317	6	ABU00733	Abu00733 S. pneumo
433	36.5	42.9	317	2	AAW61195	Aaw61195 Rhodococc	506	36	42.4	317	6	ABU097196	Abu097196 Enzyme po
434	36.5	42.9	317	5	ABG98023	Abg98023 Mutant ha	507	36	42.4	319	4	ABW63529	Abw63529 Drosophil
435	36.5	42.9	317	5	ABG98024	Abg98024 Mutant ha	508	36	42.4	319	7	ABM86168	Abm86168 Rice abio
436	36.5	42.9	317	5	ABG98022	Abg98022 Mutant ha	509	36	42.4	319	7	ABM86168	Abm86168 Rice abio
437	36.5	42.9	317	5	ABG98021	Abg98021 Mutant ha	510	36	42.4	340	4	AAW16182	Aam16182 Peptide #
438	36.5	42.9	352	4	ABG24376	Abg24376 Novel hum	511	36	42.4	340	4	ABW35162	Abw35162 Peptide #
439	36.5	42.9	446	7	ADFI18515	Adfi18515 Haloalkan	512	36	42.4	340	4	AAW28670	Aam28670 Peptide #
440	36.5	42.9	611	8	ADNA7817	Adna7817 Thermococ	513	36	42.4	340	4	ABB29991	Abb29991 Peptide #
441	36.5	42.9	1032	6	ABU15004	Abu15004 Protein e	514	36	42.4	340	4	ABB20596	Abb20596 Protein #
442	36	42.4	57	4	AAW18917	Aam18917 Peptide #	515	36	42.4	340	4	AAW68363	Aam68363 Human bon
443	36	42.4	57	4	ABB38054	Abb38054 Peptide #	516	36	42.4	340	4	AAW55995	Aam55995 Human bra
444	36	42.4	57	4	AAW31477	Aam31477 Peptide #	517	36	42.4	340	4	ABG50034	Abg50034 Human liv
445	36	42.4	57	4	ABW23279	Abw23279 Protein #	518	36	42.4	340	4	AAW03910	Aam03910 Peptide #
446	36	42.4	57	4	AAW71196	Aam71196 Human bon	519	36	42.4	340	5	ABG37926	Abg37926 Human pep
447	36	42.4	57	4	AAW58669	Aam58669 Human bra	520	36	42.4	340	9	AEBA0832	Aeb40832 L. pneumo
448	36	42.4	57	4	ABG52908	Abg52908 Human liv	521	36	42.4	342	9	AEBA0832	Aeb37520 L. pneumo
449	36	42.4	57	5	ABG40993	Abg40993 Human pep	522	36	42.4	344	6	ABU45865	Abu45865 Protein e
450	36	42.4	58	8	ABO58066	Abos8066 Human gen	523	36	42.4	344	8	ADK47444	Adk47444 Streptoco
451	36	42.4	73	4	AAW92715	Aag92715 C glutami	524	36	42.4	349	7	ADB70081	Adb70081 C. neofor
452	36	42.4	73	7	ADL66061	Adl66061 C. glutam	525	36	42.4	349	8	ADN17532	Adn17532 Bacterial
453	36	42.4	87	4	AAW75348	Aag75348 Human col	526	36	42.4	354	2	AAW35086	Aay35086 Chlamydia
454	36	42.4	105	6	ABU39319	Abu39319 Protein e	527	36	42.4	358	8	ADW50277	Adw50277 Oil-assoc
455	36	42.4	105	6	ABU30204	Abu30204 Protein e	528	36	42.4	359	8	ADR94011	Adr94011 Novel S.
456	36	42.4	109	6	ABU58246	Abu58246 Soybean s	529	36	42.4	359	9	AEA57881	Aea57881 Streptoco
457	36	42.4	109	7	ADJ70106	Adj70106 Human hea	530	36	42.4	362	4	ABW62882	Abw62882 Drosophil
458	36	42.4	120	8	ADY22590	Ady22590 Plant ful	531	36	42.4	364	4	ABB62683	Abb62683 Drosophil
459	36	42.4	124	5	AAU99739	Aau99739 Pyrodicti	532	36	42.4	370	7	ADW71304	Adw71304 Novel hum
460	36	42.4	124	7	ADB46052	Adb46052 CanE fusi	533	36	42.4	374	4	AAG666735	Aag666735 Actin 41.
461	36	42.4	139	6	ABU58194	Abu58194 Soybean s	534	36	42.4	375	7	ADW52744	Adw52744 Porphyrin e
462	36	42.4	161	6	ABU58195	Abu58195 Soybean s	535	36	42.4	396	6	ABU15401	Abu15401 Protein e

536	36	42.4	410	8	ADS29523	Adg29523 Bacterial	609	36	42.4	716	4	AAAG67530	Aag67530 Amino aci
537	36	42.4	415	8	ADJ50341	Adj50341 Oil-absoec	610	36	42.4	716	4	AAAM40376	Aam40376 Human pol
538	36	42.4	425	8	ADN27034	Adn27034 Bacterial	611	36	42.4	716	5	AAU83686	Aau83686 Human PRO
539	36	42.4	428	4	RAE11967	Rae11967 Bacillus	612	36	42.4	716	5	ABB84927	Abb84927 Human PRO
540	36	42.4	429	3	RAG31628	Rag31628 Arabidops	613	36	42.4	716	5	ABB95533	Abb95533 Human ang
541	36	42.4	430	8	ADQ80406	Adq80406 Human PAC	614	36	42.4	716	5	ADY31930	Ady31930 Novel hum
542	36	42.4	431	4	RAB94521	Rab94521 Human pro	615	36	42.4	716	5	ABO17871	Abol17871 Novel hum
543	36	42.4	434	6	ABU41213	Abu41213 Protein e	616	36	42.4	716	6	ABU80833	Abu80833 Human PRO
544	36	42.4	436	8	ADT58069	Adt58069 Plant pol	617	36	42.4	716	6	ABO33799	Abol33799 Novel hum
545	36	42.4	440	7	ADF05748	Adf05748 Bacterial	618	36	42.4	716	6	ABU81125	Abu81125 Human PRO
546	36	42.4	441	9	ADM99297	Adm99297 C_heteros	619	36	42.4	716	6	ABU71431	Abu71431 Human neo
547	36	42.4	444	9	AE319950	Aeb319950 L_pneumo	620	36	42.4	716	6	ABU66825	Abu66825 Human PRO
548	36	42.4	447	8	ADN73285	Adn73285 Thale cre	621	36	42.4	716	6	ABU59906	Abu59906 Novel sec
549	36	42.4	448	6	ADB08446	Adb08446 Alloioococ	622	36	42.4	716	6	ABO25096	Abol25096 Human sec
550	36	42.4	453	4	AAU87234	Aau87234 Novel cen	623	36	42.4	716	6	ABU82142	Abu82142 Novel hum
551	36	42.4	453	8	ADI54549	Adi54549 Novel hum	624	36	42.4	716	6	ABU67101	Abu67101 Human sec
552	36	42.4	453	9	AEB36556	Aeb36556 L_pneumo	625	36	42.4	716	6	ADA46031	Ada46031 Novel hum
553	36	42.4	463	8	ADT08200	Adt08200 A_acidoc	626	36	42.4	716	6	ADA76462	Ada76462 Human PRO
554	36	42.4	469	3	RAG31627	Rag31627 Arabidops	627	36	42.4	716	6	ABU72322	Abu72322 Human PRO
555	36	42.4	470	6	ABR41444	AbR41444 Human DIT	628	36	42.4	716	6	ADA19112	Ada19112 Human PRO
556	36	42.4	472	3	RAG31626	Rag31626 Arabidops	629	36	42.4	716	6	ADA61735	Ada61735 Homo sapi
557	36	42.4	472	8	ADN72935	Adn72935 Thale cre	630	36	42.4	716	6	ADB19520	Adb19520 Novel hum
558	36	42.4	473	8	ADT08204	Adt08204 A_acidoc	631	36	42.4	716	6	ADB28061	Adb28061 Human PRO
559	36	42.4	477	8	ADX66641	Adx66641 Plant ful	632	36	42.4	716	6	ADA86540	Ada86540 Novel hum
560	36	42.4	478	4	AAU40714	Aau40714 Propionib	633	36	42.4	716	6	ADB16104	Adb16104 Human PRO
561	36	42.4	481	6	ABM37233	Abm37233 Propionib	634	36	42.4	716	6	ADA47890	Ada47890 Human PRO
562	36	42.4	481	6	ADB08448	Adb08448 Alloioococ	635	36	42.4	716	6	ADA67685	Ada67685 Human PRO
563	36	42.4	492	4	ABG28689	Abg28689 Novel hum	636	36	42.4	716	6	ADB30692	Adb30692 Human PRO
564	36	42.4	493	6	ADB08450	Adb08450 Alloioococ	637	36	42.4	716	6	ADA85988	Ada85988 Novel hum
565	36	42.4	504	6	RAE33781	Rae33781 Human nuc	638	36	42.4	716	6	ADA97200	Ada97200 Human PRO
566	36	42.4	504	7	ADC31033	Adc31033 Human nov	639	36	42.4	716	6	ADA79504	Ada79504 Human PRO
567	36	42.4	507	8	ADU87528	Adu87528 Mycoplasma	640	36	42.4	716	6	ADA87643	Ada87643 Novel hum
568	36	42.4	507	8	ADU87530	Adu87530 Mycoplasma	641	36	42.4	716	6	ADB16845	Adb16845 Human PRO
569	36	42.4	512	6	ADB08452	Adb08452 Alloioococ	642	36	42.4	716	6	ADA91937	Ada91937 Novel hum
570	36	42.4	513	5	ABB84258	Abb84258 Human act	643	36	42.4	716	6	ADB15000	Adb15000 Human PRO
571	36	42.4	516	8	ADS24078	Ads24078 Bacterial	644	36	42.4	716	6	ADB18961	Adb18961 Novel hum
572	36	42.4	520	4	ABB67717	Abb67717 Drosophil	645	36	42.4	716	6	ADA94176	Ada94176 Human PRO
573	36	42.4	520	6	ABU34144	Abu34144 Protein e	646	36	42.4	716	6	ADB20072	Adb20072 Novel hum
574	36	42.4	525	4	ABG61741	Abg61741 Drosophil	647	36	42.4	716	6	ADB13384	Adb13384 Human PRO
575	36	42.4	544	9	ADV77075	Adv77075 Huntingto	648	36	42.4	716	6	ABO43404	Abol43404 Novel hum
576	36	42.4	544	9	ADV77074	Adv77074 Huntingto	649	36	42.4	716	6	ADA74638	Ada74638 Human PRO
577	36	42.4	550	4	ABG16717	Abg16717 Novel hum	650	36	42.4	716	6	ADB24871	Adb24871 Human PRO
578	36	42.4	558	6	ABR83611	AbR83611 Human atl	651	36	42.4	716	6	ADA82395	Ada82395 Human PRO
579	36	42.4	558	6	ABR82443	AbR82443 Human ARP	652	36	42.4	716	6	ADA75358	Ada75358 Human PRO
580	36	42.4	558	8	ADQ74857	Adq74857 Human and	653	36	42.4	716	6	ADA85436	Ada85436 Novel hum
581	36	42.4	558	8	ADR70103	Adr70103 Human atl	654	36	42.4	716	6	ADA84884	Ada84884 Novel hum
582	36	42.4	558	9	ADY70455	Ady70455 Human bet	655	36	42.4	716	6	ADB30140	Adb30140 Human PRO
583	36	42.4	566	6	ABM69014	Abm69014 Photorhab	656	36	42.4	716	6	ADA80668	Ada80668 Human PRO
584	36	42.4	569	4	ABB63246	Abb63246 Drosophil	657	36	42.4	716	6	ADA75910	Ada75910 Human PRO
585	36	42.4	583	6	AAG79678	Aag79678 Human ENZ	658	36	42.4	716	6	ADA47135	Ada47135 Human PRO
586	36	42.4	594	4	ABG19168	Abg19168 Novel hum	659	36	42.4	716	6	ADB25431	Adb25431 Human PRO
587	36	42.4	607	2	AAR32201	Aar32201 Full leng	660	36	42.4	716	6	ADA93607	Ada93607 Human PRO
588	36	42.4	607	6	ABU34494	Abu34494 Protein e	661	36	42.4	716	6	ADB26957	Adb26957 Human PRO
589	36	42.4	614	8	ADY08018	Ady08018 Plant ful	662	36	42.4	716	6	ADB31244	Adb31244 Human PRO
590	36	42.4	616	4	AAU28049	Aau28049 Novel hum	663	36	42.4	716	6	ABJ72450	Abj72450 Human PRO
591	36	42.4	624	4	AAAB94379	AaB94379 Human pro	664	36	42.4	716	6	ADA61172	Ada61172 Homo sapi
592	36	42.4	624	6	ABU36953	Abu36953 Protein e	665	36	42.4	716	6	ADB24319	Adb24319 Human PRO
593	36	42.4	624	7	ADC31102	Adc31102 Human nov	666	36	42.4	716	6	ADA96648	Ada96648 Human PRO
594	36	42.4	626	8	ABO58592	Abol58592 Human gen	667	36	42.4	716	6	ADA81220	Ada81220 Human PRO
595	36	42.4	663	3	ABM58946	Abm58946 Breast an	668	36	42.4	716	6	ADA96096	Ada96096 Human PRO
596	36	42.4	664	5	ABB93670	Abb93670 Herbicida	669	36	42.4	716	6	ADB26405	Adb26405 Human PRO
597	36	42.4	664	6	ABR56952	AbR56952 Human SDH	670	36	42.4	716	6	ADB21890	Adb21890 Novel hum
598	36	42.4	664	7	ADJ68905	Adj68905 Human hea	671	36	42.4	716	6	ABO34345	Abol34345 Human sec
599	36	42.4	664	7	ADJ70608	Adj70608 Human hea	672	36	42.4	716	7	ADA77669	Ada77669 Human PRO
600	36	42.4	664	7	ADJ68810	Adj68810 Human hea	673	36	42.4	716	7	ADB18409	Adb18409 Human PRO
601	36	42.4	664	7	ADJ68811	Adj68811 Human hea	674	36	42.4	716	7	ADA87092	Ada87092 Novel hum
602	36	42.4	664	8	ABM80599	Abm80599 Tumour-as	675	36	42.4	716	7	ADA88195	Ada88195 Novel hum
603	36	42.4	673	8	ADY13323	Ady13323 Plant ful	676	36	42.4	716	7	ADA46583	Ada46583 Novel hum
604	36	42.4	713	8	ADS24759	Ads24759 Bacterial	677	36	42.4	716	7	ADB28613	Adb28613 Human PRO
605	36	42.4	716	3	AAB33472	Aab33472 Human PRO	678	36	42.4	716	7	ADB29165	Adb29165 Human PRO
606	36	42.4	716	4	AAB27234	Aab27234 Human EXM	679	36	42.4	716	7	ADA77117	Ada77117 Human PRO
607	36	42.4	716	4	AAB50965	AbB50965 Human PRO	680	36	42.4	716	7	ADA88747	Ada88747 Novel hum
608	36	42.4	716	4	AAU12427	Aau12427 Human PRO	681	36	42.4	716	7	ADA97752	Ada97752 Human PRO

682	36	42.4	716	7	ADb27509	Human PRO	755	36	42.4	716	7	ADDS3206	Human PRO
683	36	42.4	716	7	ADb22442	Novel hum	756	36	42.4	716	7	ADDS3758	Novel hum
684	36	42.4	716	7	ABJ72152	Human mem	757	36	42.4	716	7	ADD37264	Human sec
685	36	42.4	716	7	ADa67133	Human PRO	758	36	42.4	716	7	ADDS1914	Human PRO
686	36	42.4	716	7	ADb22994	Human PRO	759	36	42.4	716	7	ADD02713	Human PRO
687	36	42.4	716	7	ADb23767	Human PRO	760	36	42.4	716	7	ADDS0581	Human PRO
688	36	42.4	716	7	ADa92489	Novel hum	761	36	42.4	716	7	ADD02147	Human PRO
689	36	42.4	716	7	ADb15552	Human PRO	762	36	42.4	716	7	ADDS4329	Novel hum
690	36	42.4	716	7	ADb83680	Novel hum	763	36	42.4	716	7	ADDS0335	Human PRO
691	36	42.4	716	7	ADb80786	Novel hum	764	36	42.4	716	7	ADDS1346	Novel hum
692	36	42.4	716	7	ADb73327	Novel hum	765	36	42.4	716	7	ADD92646	Human PRO
693	36	42.4	716	7	ADb38804	Novel hum	766	36	42.4	716	7	ADD91542	Human PRO
694	36	42.4	716	7	ADb78409	Novel hum	767	36	42.4	716	7	ADe04156	Human PRO
695	36	42.4	716	7	ADb38252	Novel hum	768	36	42.4	716	7	ADe32453	Novel hum
696	36	42.4	716	7	ADb66724	Novel hum	769	36	42.4	716	7	ADe22385	Human PRO
697	36	42.4	716	7	ADb85057	Human PRO	770	36	42.4	716	7	ADd79609	Human PRO
698	36	42.4	716	7	ADb89804	Human PRO	771	36	42.4	716	7	ADe42145	Human PRO
699	36	42.4	716	7	ADb90536	Human PRO	772	36	42.4	716	7	ADe17962	Human PRO
700	36	42.4	716	7	ADb39637	Novel hum	773	36	42.4	716	7	ADD92094	Human PRO
701	36	42.4	716	7	ADb78163	Novel hum	774	36	42.4	716	7	ADe33557	Novel hum
702	36	42.4	716	7	ADb87229	Human PRO	775	36	42.4	716	7	ADe34109	Novel hum
703	36	42.4	716	7	ADb84811	Human PRO	776	36	42.4	716	7	ADd80161	Human PRO
704	36	42.4	716	7	ADb47260	Novel hum	777	36	42.4	716	7	ADD93198	Human PRO
705	36	42.4	716	7	ADb83926	Novel hum	778	36	42.4	716	7	ADe19618	Human PRO
706	36	42.4	716	7	ADb86867	Human PRO	779	36	42.4	716	7	ADe19066	Human PRO
707	36	42.4	716	7	ADb73081	Novel hum	780	36	42.4	716	7	ADe43262	Human PRO
708	36	42.4	716	7	ADb77472	Novel hum	781	36	42.4	716	7	ADd96051	Human PRO
709	36	42.4	716	7	ADb34629	Human PRO	782	36	42.4	716	7	ADe22937	Human PRO
710	36	42.4	716	7	ADb35733	Human PRO	783	36	42.4	716	7	ADD79055	Human PRO
711	36	42.4	716	7	ADb34077	Human PRO	784	36	42.4	716	7	ADe33005	Novel hum
712	36	42.4	716	7	ADb35181	Human PRO	785	36	42.4	716	7	ADe42697	Human PRO
713	36	42.4	716	7	ADb36285	Human PRO	786	36	42.4	716	7	ADd80713	Human PRO
714	36	42.4	716	7	ADb46680	Novel hum	787	36	42.4	716	7	ADd89741	Human PRO
715	36	42.4	716	7	ADc36919	Human PRO	788	36	42.4	716	7	ADe41025	Human PRO
716	36	42.4	716	7	ADc21909	Human PRO	789	36	42.4	716	7	ADe04824	Human PRO
717	36	42.4	716	7	ADc50553	Novel hum	790	36	42.4	716	7	ADe92953	Human PRO
718	36	42.4	716	7	ADc72100	Novel hum	791	36	42.4	716	7	ADG21662	Novel hum
719	36	42.4	716	7	ADc60079	Novel hum	792	36	42.4	716	7	ADG23303	Novel hum
720	36	42.4	716	7	ADc49940	Novel hum	793	36	42.4	716	7	ADf97638	Human PRO
721	36	42.4	716	7	ADc49139	Novel hum	794	36	42.4	716	7	ADG80702	Human PRO
722	36	42.4	716	7	ADc49656	Novel hum	795	36	42.4	716	7	ADg80150	Human PRO
723	36	42.4	716	7	ADc47517	Novel hum	796	36	42.4	716	7	ADh55442	Novel hum
724	36	42.4	716	7	ADc53086	Novel hum	797	36	42.4	716	7	ADh55994	Novel hum
725	36	42.4	716	7	ADc57440	Novel hum	798	36	42.4	716	7	ADi64213	Novel hum
726	36	42.4	716	7	ADc60631	Novel hum	799	36	42.4	716	7	ADi65162	Novel hum
727	36	42.4	716	7	ADc51106	Novel hum	800	36	42.4	716	7	ADi63661	Novel hum
728	36	42.4	716	7	ADc65633	Human PRO	801	36	42.4	716	7	ADh82075	Novel hum
729	36	42.4	716	7	ADc54731	Novel hum	802	36	42.4	716	7	ADh81523	Novel hum
730	36	42.4	716	7	ADc53692	Novel hum	803	36	42.4	716	7	ADm82692	Novel hum
731	36	42.4	716	7	ADc59215	Novel hum	804	36	42.4	716	7	ADn16091	Novel hum
732	36	42.4	716	7	ADc56093	Novel hum	805	36	42.4	716	7	ADn16720	Novel hum
733	36	42.4	716	7	ADc58663	Novel hum	806	36	42.4	716	7	ADn15539	Novel hum
734	36	42.4	716	7	ADc47262	Novel hum	807	36	42.4	716	7	ADn14987	Novel hum
735	36	42.4	716	7	ADD03337	Novel hum	808	36	42.4	716	8	ADc48893	Novel hum
736	36	42.4	716	7	ADc90329	Novel hum	809	36	42.4	716	8	ADc81249	Novel hum
737	36	42.4	716	7	ADc69748	Human PRO	810	36	42.4	716	8	ADe21064	Novel hum
738	36	42.4	716	7	ADc48637	Human PRO	811	36	42.4	716	8	ADe05908	Human PRO
739	36	42.4	716	7	ADD10166	Human PRO	812	36	42.4	716	8	ADd76697	Human PRO
740	36	42.4	716	7	ADc78137	Novel hum	813	36	42.4	716	8	ADD75137	Human PRO
741	36	42.4	716	7	ADD04741	Novel hum	814	36	42.4	716	8	ADD75883	Novel hum
742	36	42.4	716	7	ADD06372	Novel hum	815	36	42.4	716	8	ADd85115	Novel hum
743	36	42.4	716	7	ADc80697	Novel hum	816	36	42.4	716	8	ADd86941	Novel hum
744	36	42.4	716	7	ADD11204	Human PRO	817	36	42.4	716	8	ADd820818	Novel hum
745	36	42.4	716	7	ADD10511	Human sec	818	36	42.4	716	8	ADe39115	Novel hum
746	36	42.4	716	7	ADc48085	Human PRO	819	36	42.4	716	8	ADd88061	Human PRO
747	36	42.4	716	7	ADc77891	Novel hum	820	36	42.4	716	8	ADd86465	Human PRO
748	36	42.4	716	7	ADc80145	Novel hum	821	36	42.4	716	8	ADe05662	Human PRO
749	36	42.4	716	7	ADD11471	Human sec	822	36	42.4	716	8	ADd73647	Human PRO
750	36	42.4	716	7	ADD09614	Human PRO	823	36	42.4	716	8	ADd75913	Human PRO
751	36	42.4	716	7	ADd50854	Novel hum	824	36	42.4	716	8	ADd78487	Novel hum
752	36	42.4	716	7	ADD41327	Novel hum	825	36	42.4	716	8	ADe41472	Human sec
753	36	42.4	716	7	ADD52466	Human PRO	826	36	42.4	716	8	ADe23489	Human PRO
754	36	42.4	716	7	ADDS1100	Novel hum	827	36	42.4	716	8	ADe21310	Novel hum



828	36	42.4	716	8	ADD77425	Novel hum	Ad777425	Novel hum	901	36	42.4	716	8	ADG53733	Novel hum
829	36	42.4	716	8	ADE20572	Novel hum	Ade20572	Novel hum	902	36	42.4	716	8	ADG71619	Novel hum
830	36	42.4	716	8	ADD75637	Human PRO	Ad75637	Human PRO	903	36	42.4	716	8	ADG81806	Human PRO
831	36	42.4	716	8	ADD74153	Human PRO	Ad74153	Human PRO	904	36	42.4	716	8	ADH30768	Human PRO
832	36	42.4	716	8	ADD74399	Human PRO	Ad74399	Human PRO	905	36	42.4	716	8	ADH12135	Novel hum
833	36	42.4	716	8	ADD76129	Novel hum	Ad76129	Novel hum	906	36	42.4	716	8	ADG52557	Novel hum
834	36	42.4	716	8	ADD85621	Novel hum	Ad85621	Novel hum	907	36	42.4	716	8	ADG54285	Novel hum
835	36	42.4	716	8	ADE24041	Human PRO	Ade24041	Human PRO	908	36	42.4	716	8	ADG81254	Human PRO
836	36	42.4	716	8	ADE24684	Human PRO	Ade24684	Human PRO	909	36	42.4	716	8	ADG56493	Novel hum
837	36	42.4	716	8	ADD87509	Human PRO	Ad87509	Human PRO	910	36	42.4	716	8	ADH12759	Novel hum
838	36	42.4	716	8	ADE05170	Human PRO	Ade05170	Human PRO	911	36	42.4	716	8	ADG61605	Novel hum
839	36	42.4	716	8	ADD75383	Human PRO	Ad75383	Human PRO	912	36	42.4	716	8	ADH28692	Human PRO
840	36	42.4	716	8	ADD76927	Novel hum	Ad76927	Novel hum	913	36	42.4	716	8	ADG54837	Novel hum
841	36	42.4	716	8	ADD86695	Novel hum	Ad86695	Novel hum	914	36	42.4	716	8	ADG59877	Novel hum
842	36	42.4	716	8	ADE89375	Human PRO	Ade89375	Human PRO	915	36	42.4	716	8	ADG78872	Human can
843	36	42.4	716	8	ADD78163	Novel hum	Ad78163	Novel hum	916	36	42.4	716	8	ADH43655	Human PRO
844	36	42.4	716	8	ADE18514	Human PRO	Ade18514	Human PRO	917	36	42.4	716	8	ADG34186	Novel hum
845	36	42.4	716	8	ADE88823	Human PRO	Ade88823	Human PRO	918	36	42.4	716	8	ADH13101	Human PRO
846	36	42.4	716	8	ADD77671	Novel hum	Ad77671	Novel hum	919	36	42.4	716	8	ADI13656	Human PRO
847	36	42.4	716	8	ADD77917	Novel hum	Ad77917	Novel hum	920	36	42.4	716	8	ADH69750	Human PRO
848	36	42.4	716	8	ADD85375	Novel hum	Ad85375	Novel hum	921	36	42.4	716	8	ADG10044	Novel hum
849	36	42.4	716	8	ADD73907	Human PRO	Ad73907	Human PRO	922	36	42.4	716	8	ADI15515	Novel hum
850	36	42.4	716	8	ADD74645	Human PRO	Ad74645	Human PRO	923	36	42.4	716	8	ADG09392	Novel hum
851	36	42.4	716	8	ADD77173	Novel hum	Ad77173	Novel hum	924	36	42.4	716	8	ADI14847	Novel hum
852	36	42.4	716	8	ADD85867	Novel hum	Ad85867	Novel hum	925	36	42.4	716	8	ADI29911	Novel hum
853	36	42.4	716	8	ADE05416	Human PRO	Ade05416	Human PRO	926	36	42.4	716	8	ADI18442	Novel hum
854	36	42.4	716	8	ADD74891	Human PRO	Ad74891	Human PRO	927	36	42.4	716	8	ADM27308	Novel hum
855	36	42.4	716	8	ADE94843	Human PRO	Ade94843	Human PRO	928	36	42.4	716	8	ADJ63723	Novel hum
856	36	42.4	716	8	ADE91254	Human PRO	Ade91254	Human PRO	929	36	42.4	716	8	ADJ77618	Human PRO
857	36	42.4	716	8	ADE95395	Human PRO	Ade95395	Human PRO	930	36	42.4	716	8	ADK83000	Human PRO
858	36	42.4	716	8	ADE93505	Human PRO	Ade93505	Human PRO	931	36	42.4	716	8	ADK66666	Human PRO
859	36	42.4	716	8	ADF35086	Human PRO	Adf35086	Human PRO	932	36	42.4	716	8	ADJ65740	Human PRO
860	36	42.4	716	8	ADE92401	Novel hum	Ade92401	Novel hum	933	36	42.4	716	8	ADM27876	Human PRO
861	36	42.4	716	8	ADE90702	Human PRO	Ade90702	Human PRO	934	36	42.4	716	8	ADM42600	Human PRO
862	36	42.4	716	8	ADE91849	Novel hum	Ade91849	Novel hum	935	36	42.4	716	8	ADM28462	Human PRO
863	36	42.4	716	8	ADG05703	Novel hum	Adg05703	Novel hum	936	36	42.4	716	8	ADI95944	Human PRO
864	36	42.4	716	8	ADG27257	Human PRO	Adg27257	Human PRO	937	36	42.4	716	8	ADI96496	Novel hum
865	36	42.4	716	8	ADG02428	Human PRO	Adg02428	Human PRO	938	36	42.4	716	8	ADS32448	Novel hum
866	36	42.4	716	8	ADG22214	Novel hum	Adg22214	Novel hum	939	36	42.4	716	8	ADT03432	Human PRO
867	36	42.4	716	8	ADG20284	Human PRO	Adg20284	Human PRO	940	36	42.4	716	9	ADZ03483	Human sec
868	36	42.4	716	8	ADF98190	Human PRO	Adf98190	Human PRO	941	36	42.4	716	9	AEA54850	Human tum
869	36	42.4	716	8	ADG24407	Novel hum	Adg24407	Novel hum	942	36	42.4	716	9	AEI14229	Cancer ce
870	36	42.4	716	8	ADF98761	Human PRO	Adf98761	Human PRO	943	36	42.4	718	2	AEY28883	Human Ext
871	36	42.4	716	8	ADG03592	Human PRO	Adg03592	Human PRO	944	36	42.4	718	6	ABR61847	Exotoxins
872	36	42.4	716	8	ADF99313	Human PRO	Adf99313	Human PRO	945	36	42.4	718	8	ADQ17892	Human sof
873	36	42.4	716	8	ADG16898	Human PRO	Adg16898	Human PRO	946	36	42.4	718	8	ADUI8070	Human can
874	36	42.4	716	8	ADG05357	Human PRO	Adg05357	Human PRO	947	36	42.4	728	2	AAW44851	Hereditar
875	36	42.4	716	8	ADG19624	Human PRO	Adg19624	Human PRO	948	36	42.4	728	6	ABG74453	Human exo
876	36	42.4	716	8	ADG11320	Novel hum	Adg11320	Novel hum	949	36	42.4	728	7	ADJ71171	Human hea
877	36	42.4	716	8	ADG13461	Human PRO	Adg13461	Human PRO	950	36	42.4	730	6	AAO23111	LRRN1 'hu
878	36	42.4	716	8	ADG08518	Novel hum	Adg08518	Novel hum	951	36	42.4	730	7	ADB75382	Prostate
879	36	42.4	716	8	ADG15688	Human PRO	Adg15688	Human PRO	952	36	42.4	730	8	ADR66436	Human pro
880	36	42.4	716	8	ADG12099	Novel hum	Adg12099	Novel hum	953	36	42.4	730	8	ADR66094	Human pro
881	36	42.4	716	8	ADF97086	Human PRO	Adf97086	Human PRO	954	36	42.4	734	3	AAG29282	Arabidops
882	36	42.4	716	8	ADG06271	Human PRO	Adg06271	Human PRO	955	36	42.4	736	3	AAG29281	Arabidops
883	36	42.4	716	8	ADG23855	Novel hum	Adg23855	Novel hum	956	36	42.4	750	3	AAB26599	Synechocy
884	36	42.4	716	8	ADG04144	Human PRO	Adg04144	Human PRO	957	36	42.4	756	4	ABG13914	Novel hum
885	36	42.4	716	8	ADG25045	Novel hum	Adg25045	Novel hum	958	36	42.4	760	4	ABB65550	Drosophil
886	36	42.4	716	8	ADF94656	Novel hum	Adf94656	Novel hum	959	36	42.4	763	4	AAU41222	Propionib
887	36	42.4	716	8	ADG07342	Novel hum	Adg07342	Novel hum	960	36	42.4	763	6	ABM17741	Propionib
888	36	42.4	716	8	ADG07894	Novel hum	Adg07894	Novel hum	961	36	42.4	782	6	ABU49972	Protein e
889	36	42.4	716	8	ADG06752	Human PRO	Adg06752	Human PRO	962	36	42.4	805	8	ADM94430	Human pho
890	36	42.4	716	8	ADG55389	Novel hum	Adg55389	Novel hum	963	36	42.4	812	8	ADG78871	Human can
891	36	42.4	716	8	ADG61053	Novel hum	Adg61053	Novel hum	964	36	42.4	816	5	ABP73898	Canida a
892	36	42.4	716	8	ADG62157	Novel hum	Adg62157	Novel hum	965	36	42.4	824	2	AAR04575	Derived a
893	36	42.4	716	8	ADG82358	Human PRO	Adg82358	Human PRO	966	36	42.4	825	2	AAH11254	Human IL-
894	36	42.4	716	8	ADG57597	Novel hum	Adg57597	Novel hum	967	36	42.4	825	2	AAH13499	Human int
895	36	42.4	716	8	ADG57045	Novel hum	Adg57045	Novel hum	968	36	42.4	825	2	AAW48314	Human int
896	36	42.4	716	8	ADG55941	Novel hum	Adg55941	Novel hum	969	36	42.4	825	2	AAW60668	Human int
897	36	42.4	716	8	ADG58701	Novel hum	Adg58701	Novel hum	970	36	42.4	825	2	AAW73470	Human int
898	36	42.4	716	8	ADG71067	Novel hum	Adg71067	Novel hum	971	36	42.4	825	2	AAW93004	Human IL-
899	36	42.4	716	8	ADH39096	Novel hum	Adh39096	Novel hum	972	36	42.4	825	2	AAW23920	Human EST
900	36	42.4	716	8	ADG58149	Novel hum	Adg58149	Novel hum	973	36	42.4	825	4	AAU07612	Human int

974 36 42.4 825 4 AAM38752 Human pol  
 975 36 42.4 825 4 AAB66970 Human IL4  
 976 36 42.4 825 5 AAU77212 Human int  
 977 36 42.4 825 5 AAU77214 Human int  
 978 36 42.4 825 5 AAU78090 Human int  
 979 36 42.4 825 5 ABB07227 Human int  
 980 36 42.4 825 7 ADE64017 Human pro  
 981 36 42.4 825 7 ADF17833 Human IL-  
 982 36 42.4 825 7 ADM31001 Human int  
 983 36 42.4 825 7 ADN95726 Human BEC  
 984 36 42.4 825 7 ADP65149 Human int  
 985 36 42.4 825 7 ADP65144 Human int  
 986 36 42.4 825 7 ABM85533 Human pro  
 987 36 42.4 825 8 ADF78073 Human int  
 988 36 42.4 825 8 ADL82825 Human pro  
 989 36 42.4 825 8 ADN03942 Antipsori  
 990 36 42.4 825 8 ADP12992 Protein e  
 991 36 42.4 825 8 ADO19357 Human pro  
 992 36 42.4 825 8 ADQ39449 Human myo  
 993 36 42.4 825 8 ADV15273 Human ost  
 994 36 42.4 825 9 ADX05058 Human IL4  
 995 36 42.4 825 9 ADY15032 PRO polyp  
 996 36 42.4 825 9 ADY19482 PRO polyp  
 997 36 42.4 825 9 ADZ13003 Human can  
 998 36 42.4 825 9 ADZ70502 Human pro  
 999 36 42.4 825 9 AEA08877 Human IL-  
 1000 36 42.4 825 9 AEB04226 Human IL-

## ALIGNMENTS

RESULT 1  
 AAR43698  
 ID AAR43698 standard; peptide; 14 AA.  
 XX  
 AC AAR43698;  
 XX  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 20-MAY-1994 (first entry)  
 XX  
 DE GT subsequence.  
 XX  
 KW GT; glucosyltransferase; vaccine; T-cell; B-cell; reaction;  
 KW immunoresponse; peptidyl core matrix; dental caries; diptheria; tetanus;  
 KW measles; polio.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9322341-A1.  
 XX  
 PD 11-NOV-1993.  
 XX  
 PF 30-APR-1993; 93WO-US004094.  
 XX  
 PR 01-MAY-1992; 92US-00877295.  
 XX  
 PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.  
 XX  
 PI Taubman MA, Smith DJ;  
 XX  
 DR WPI; 1993-368721/46.  
 XX  
 PT Synthetic glucosyl-transferase peptide(s) - provoke T- and B-cell  
 PT reactions in mammals, and are effective against dental caries.  
 XX  
 PS Claim 3; Page 24; 38pp; English.  
 XX  
 CC The sequences (AAR43694-98) are subsequences from GT, they provoke T- and  
 CC B-cell responses in mammals. The can be used to produce a vaccine  
 CC comprising 2 of these sequences attached to a peptidyl core matrix. They  
 CC are useful in producing T-cell responses to the GT enzyme causing a  
 CC reduction of colonisation or accumulation of mutans streptococcal strains

CC in mammals. The vaccines can be used in preventing dental caries.  
 CC (Updated on 25-MAR-2003 to correct PN field.)  
 XX  
 SQ Sequence 14 AA;  
 Query Match 100.0%; Score 85; DB 2; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-07;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QWNGSESEKPYDDHL 14  
 Db 1 QWNGSESEKPYDDHL 14  
 RESULT 2  
 AAW34160  
 ID AAW34160 standard; peptide; 14 AA.  
 XX  
 AC AAW34160;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 18-FEB-1998 (first entry)  
 XX  
 DE GTF antigenic peptide #5.  
 XX  
 KW Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;  
 KW surface domain; glucan-binding domain; mutans streptococcal strain;  
 KW immune response; glucan-binding activity; dental caries prevention.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PN US5686075-A.  
 XX  
 PD 11-NOV-1997.  
 XX  
 PF 30-APR-1993; 93US-00057162.  
 XX  
 PR 01-MAY-1992; 92US-00877295.  
 XX  
 PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX  
 DR WPI; 1997-558089/51.  
 XX  
 PT Immunogenic compositions containing streptococcal glucosyl-transferase  
 PT peptide(s) - used for provoking immune response to streptococcal  
 PT glucosyl-transferase for preventing dental caries.  
 XX  
 PS Claim 6; Col 14; 11pp; English.  
 XX  
 CC AAW34156-W34160 represent immunogenic fragments of the Streptomyces  
 CC mutans glucosyltransferase (GTF) enzyme. AAW34157 and AAW34158 are from  
 CC the catalytic domain of GTF, while AAW34160 is from the GTF surface  
 CC domain. AAW34156 and AAW34159 are from the glucan-binding domain of GTF.  
 CC These sequences can all be used in the immunogenic composition of the  
 CC invention. The composition of the invention can alternatively comprise  
 CC one or more of these sequences linked to a lysine core matrix (see  
 CC AAW34161-W34165). A composition comprising one of these sequences can be  
 CC administered to a mammal to raise an immune response, in a method for  
 CC interfering with the enzymatic activity of streptococcal  
 CC glucosyltransferase in a mammal. The immune response results in reduction  
 CC of the colonisation or accumulation of mutans streptococcal strains in  
 CC the mammal. Compositions containing AAW34156 specifically interfere with  
 CC the glucan-binding activity of the streptococcal glucosyltransferase. The  
 CC peptides can also be used in vaccines for preventing dental caries in  
 CC mammals. (Updated on 25-MAR-2003 to correct PF field.)  
 XX  
 SQ Sequence 14 AA;  
 Query Match 100.0%; Score 85; DB 2; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-07;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWNGSEKPYDDHL 14  
 ID AAY43339 standard; peptide; 14 AA.  
 XX |||||  
 Db 1 QWNGSEKPYDDHL 14

RESULT 3  
 AAY43339  
 ID AAY43339 standard; peptide; 14 AA.  
 XX |||||  
 AC AAY43339;  
 XX |||||  
 DT 26-JAN-2000 (first entry)  
 XX |||||

XX GTF antigenic epitope.  
 XX |||||

XX GTF; antigenic epitope; glucosyltransferase; immunogenic epitope; teeth;  
 KW immunogenic composition; streptococcal glucan; antibody response;  
 KW cariogenic bacteria; mutans streptococci; colonisation; caries;  
 KW dental caries; immunisation; therapy.  
 OS Synthetic.  
 XX |||||

XX WO9952548-A2.  
 PN |||||

XX 21-OCT-1999.  
 PD |||||

XX 09-APR-1999; 99WO-US007828.  
 PF |||||

XX 10-APR-1998; 98US-0081315P.  
 PR |||||

XX (LEES/) LEES A.  
 PA |||||

XX (TAUB/) TAUBMAN M A.  
 PA |||||

XX (SMIT/) SMITH D J.  
 PA |||||

XX Lees A, Taubman MA, Smith DJ;  
 PI |||||

XX WPI; 1999-620289/53.  
 DR |||||

XX Immunogenic compositions for control of dental caries, based on  
 PT Streptococcus mutans components, particularly for vaccination of infants.

XX Claim 11; Page 44; 54pp; English.  
 PS |||||

XX This sequence represents a glucosyltransferase (GTF) immunogenic epitope,  
 CC and can be used in the immunogenic composition of the invention. The  
 CC immunogenic composition (A) comprises at least one streptococcal glucan  
 CC (SG) and at least one component (I) covalently coupled, (in)directly, to  
 CC SG. (A) elicits antibodies (Ab) to both SG and (I). (A), and related  
 CC compositions, are used to induce an antibody response to cariogenic  
 CC bacteria, collectively called "mutans streptococci" (i.e. any of  
 CC Streptococcus mutans, S. rattus, S. cricetus or S. sobrinus), especially  
 CC to prevent colonisation, to reduce numbers of bacteria or to reduce  
 CC incidence of caries, in humans or other animals. The composition can  
 CC therefore be used for the treatment or prevention of caries (particularly  
 CC dental caries). Abs raised against (A) can be used: (a) to detect SG, and  
 CC (b) for passive immunisation, e.g. applied directly to the surface of  
 CC teeth. SG is a poorly immunogenic type 2 T-cell independent antigen but  
 CC when coupled to (I) it produces a conjugate which induces significant  
 CC immunity to both components, resulting in immunological memory and long-  
 CC lasting antibody production against SG, in adults and children  
 XX Sequence 14 AA;  
 SQ |||||

Query Match 100.0%; Score 85; DB 2; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-07;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWNGSEKPYDDHL 14  
 ID 1 QWNGSEKPYDDHL 14

RESULT 4  
 ADD93648  
 ID ADD93648 standard; peptide; 14 AA.  
 XX |||||  
 AC ADD93648;  
 XX |||||

DT 29-JAN-2004 (first entry)  
 XX |||||

XX Streptococcus mutans glucosyltransferase surface domain peptide.  
 DE |||||

XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.  
 KW |||||

OS Streptococcus mutans.  
 XX |||||

XX WO2003075845-A2.  
 PN |||||

XX 18-SEP-2003.  
 PD |||||

XX 07-MAR-2003; 2003WO-US006962.  
 PF |||||

XX 07-MAR-2002; 2002US-0363209P.  
 PR |||||

XX 08-AUG-2002; 2002US-0402483P.  
 PR |||||

XX (FORS-) FORSYTH INST.  
 PA |||||

XX Smith DJ, Taubman MA;  
 PI |||||

XX WPI; 2003-845091/78.  
 DR |||||

XX Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.  
 PT |||||

XX Claim 20; Page 17; 49pp; English.  
 PS |||||

XX The present sequence is that of a surface domain peptide fragment of  
 CC Streptococcus mutans glucosyltransferase. The peptide can be used in  
 CC immunogenic compositions and subunit vaccines for dental caries. These  
 CC compositions comprise a major histocompatibility complex (MHC) class II  
 CC protein-binding peptide from Streptococcus mutans glucan binding protein  
 CC -B (GbpB) covalently linked with a peptide subunit of a streptococcal  
 CC glucosyltransferase. The compositions are used in a claimed method of  
 CC eliciting production of an antibody in a mammal. Dispitopic or  
 CC multiepitopic polypeptides can be prepared synthetically or by  
 CC recombinant DNA technology. Antibodies raised against MHC class II  
 CC binding fragments of GbpB can be used in passive immunisation.  
 XX Sequence 14 AA;  
 SQ |||||

Query Match 100.0%; Score 85; DB 7; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 7.8e-07;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWNGSEKPYDDHL 14  
 Db |||||  
 Db 1 QWNGSEKPYDDHL 14

RESULT 5  
 ADX37271  
 ID ADX37271 standard; peptide; 14 AA.  
 XX |||||  
 AC ADX37271;  
 XX |||||

DT 21-APR-2005 (first entry)  
 XX |||||

XX Streptococcus mutant glucan binding protein B peptide #28.  
 DE |||||

XX immunogenicity; immune stimulation; glucan binding protein-B;  
 KW microparticle; major histocompatibility complex; tooth disease.  
 KW |||||  
 OS Streptococcus mutans.  
 XX |||||



Qy 1 QWNGSEKPYDDHL 14  
 Db 2 QWNGSEKPYDDHL 15

RESULT 8  
 ADV68554  
 ID ADV68554 standard; peptide; 15 AA.  
 XX  
 AC ADV68554;  
 DT 24-FEB-2005 (first entry)  
 XX Glucosyltransferase (GTF) control peptide, MAC.  
 DE  
 XX Immunogenicity; vaccine; glucosyltransferase; GTF; tooth disease;  
 KW mouth disease; antibacterial; MAC peptide.  
 XX  
 OS Streptococcus sobrinus.  
 XX Streptococcus downei.  
 XX US6827936-B1.  
 XX  
 XX 07-DEC-2004.  
 XX  
 XX 12-APR-1999; 99US-00290049.  
 XX  
 XX 13-APR-1998; 98US-0081550P.  
 PR 08-JAN-1999; 99US-0115142P.  
 XX (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.  
 PA Smith DJ, Taubman MA;  
 PI WPI; 2005-009976/01.  
 XX  
 DR New immunogenic composition comprises a peptide corresponding to a  
 PT subunit of glucosyltransferase, useful as glucosyltransferase subunit  
 PT vaccine for preventing dental caries.  
 XX  
 XX Example 1; SEQ ID NO 4; 18pp; English.  
 XX  
 CC The present invention relates to a novel immunogenic composition which  
 CC comprises a peptide corresponding to a subunit of glucosyltransferase  
 CC (GTF) enzyme. The immunogenic composition is useful as a GTF subunit  
 CC vaccine for preventing dental caries. The present sequence is a GTF  
 CC control peptide, MAC. This sequence is found in Streptococcus sobrinus  
 CC and Streptococcus downei.  
 XX  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 85; DB 9; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-07; Indels 0; Gaps 0;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QWNGSEKPYDDHL 14  
 Db 2 QWNGSEKPYDDHL 15

RESULT 9  
 AAW34165  
 ID AAW34165 standard; peptide; 16 AA.  
 XX  
 AC AAW34165;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 18-FEB-1998 (first entry)  
 XX  
 XX GTF antigenic peptide #5 linked to polylysine core.  
 DE  
 XX Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;  
 KW surface domain; glucan-binding domain; mutans streptococcal strain;

immune response; glucan-binding activity; dental caries prevention.  
 Synthetic.  
 OS Streptococcus mutans.  
 XX  
 XX Key Location/Qualifiers  
 FT Peptide 1..14  
 FT /label= GTF antigenic peptide #5 (see AAW34160)  
 FT /note= "attached to the dendritic polylysine core via the  
 FT alpha-amino group of Lys(15); a second copy of the  
 FT antigenic 14-mer is linked to Lys(15) via the omega amino  
 FT group"  
 FT 15  
 FT Modified-site  
 FT /note= "Lys(15) is linked to one copy of the antigenic  
 FT peptide through the alpha-amino group, and to a second  
 FT copy of the peptide (not shown) via the omega amino  
 FT group"  
 FT 16  
 FT Modified-site  
 FT /note= "the alpha amino acid group of Lys(16) forms a  
 FT peptide linkage with the carboxyl amino group of Lys(15);  
 FT the omega amino group of Lys(16) forms a peptide bond  
 FT with a second Lys residue analogous to Lys(15)"  
 XX  
 XX US5686075-A.  
 XX  
 XX 11-NOV-1997.  
 PD  
 XX 30-APR-1993; 93US-00057162.  
 PF  
 XX 01-MAY-1992; 92US-00877295.  
 PR  
 XX (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.  
 PA Smith DJ, Taubman MA;  
 PI WPI; 1997-558089/51.  
 XX  
 DR Immunogenic compositions containing streptococcal glucosyltransferase  
 PT peptide(s) - used for provoking immune response to streptococcal  
 PT glucosyltransferase for preventing dental caries.  
 XX  
 XX Claim 12; Col; 11pp; English.  
 PS  
 CC AAW34161-W34165 represent the Streptomyces mutans glycosyltransferase  
 CC (GTF) enzyme immunogenic fragments shown in AAW34156-W34160 linked to a  
 CC polylysine core. AAW34157 and AAW34158 are from the catalytic domain of  
 CC GTF, while AAW34160 is from the GTF surface domain. AAW34156 and AAW34159  
 CC are from the glucan-binding domain of GTF. These sequences, and the  
 CC immunogenic fragments shown in AAW34156-W34160 can all be used in the  
 CC immunogenic composition of the invention. A composition comprising one of  
 CC these sequences can be administered to a mammal to raise an immune  
 CC response, in a method for interfering with the enzymatic activity of  
 CC streptococcal glucosyltransferase in a mammal. The immune response  
 CC results in reduction of the colonisation or accumulation of mutans  
 CC streptococcal strains in the mammal. Compositions containing AAW34156  
 CC specifically interfere with the glucan-binding activity of the  
 CC streptococcal glucosyltransferase. The peptides can also be used in  
 CC vaccines for preventing dental caries in mammals. (Updated on 25-MAR-2003  
 CC to correct Pf field.)  
 XX  
 SQ Sequence 16 AA;  
 Query Match 100.0%; Score 85; DB 2; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 9.1e-07;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QWNGSEKPYDDHL 14  
 Db 1 QWNGSEKPYDDHL 14

RESULT 10  
 ABB98615

```

ID ABB98615 standard; peptide; 17 AA.
XX
AC ABB98615;
XX
DT 29-AUG-2003 (revised)
DT 14-JAN-2003 (first entry)
XX
DE Dextrane-saccharase, DSR-E, catalytic domain conserved peptide GTFI #1.
XX
DE Dextran saccharase; enzyme; cytostatic; dermatological; antiseborrheic;
KW DSR-E; glycosyl transferase; dextran; prebiotic; pharmaceutical;
KW microflora regulation; intestinal transit; mineral assimilation;
KW colon cancer; acne; dandruff; body odour.
XX
OS Leuconostoc mesenteroides; NRRL B-1299.
XX
PN FR2822163-A1.
XX
PD 20-SEP-2002.
XX
PF 19-DEC-2001; 2001FR-00016495.
XX
PR 16-MAR-2001; 2001FR-00003631.
XX
PA (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI Bozonnet SAM, Renaud SMMC, Willemot RML, Monean PEF;
XX
DR WPI; 2002-715213/78.
XX
XX New glycosyl transferase enzymes, containing glucan bonding and catalytic
PT domains and producing alpha-(1-2) branched dextrans, useful in probiotic,
PT pharmaceutical or cosmetic compositions.
XX
PS Example 2; Fig 5; 82pp; French.
XX
XX The present invention relates to a novel dextran saccharase, DSR-E
CC (ABB98574), from Leuconostoc mesenteroides NRRL B-1299. The saccharase
CC has glycosyl transferase activity suitable for producing dextrans having
CC alpha(1-2) branches from sucrose, p-nitrophenyl-alpha-D-glucopyranoside,
CC alpha-fluoroglucose, alpha-D-glucopyranoside-alpha-D-sorbypyranoside or
CC alpha-D-galactopyranosyl-sucrose. The dextran saccharase is useful in
CC prebiotic, pharmaceutical or cosmetic compositions. The dextrans and
CC related compounds having alpha(1-2) bonds, produced using DSR-E, may be
CC involved in signalling/cellular recognition processes in vivo
CC (specifically in regulation of microflora in the intestines or on the
CC skin); and are potentially useful for improving intestinal transit,
CC increasing assimilation of minerals (e.g. calcium and/or magnesium),
CC preventing cancer of the colon and combating skin problems such as acne,
CC dandruff and body odour. The present sequence is a conserved peptide
CC sequence from the catalytic domain of the DSR-E of the invention.
CC (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 17 AA;

Query Match 100.0%; Score 85; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 9.7e-07;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDHL 14
Db |||||
2 QWNGESEKPYDDHL 15

RESULT 11
ADD93657
ID ADD93657 standard; protein; 1590 AA.
XX
AC ADD93657;
XX
XX 29-JAN-2004 (first entry)
DT
XX
DE Streptococcus sobrinus glucosyltransferase-I.

XX
XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus sobrinus.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
PA (FORS-) FORSYTH INST.
XX
PI Smith DJ, Taubman MA;
XX
DR WPI; 2003-845091/78.
XX
PT Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
PS Claim 16; Page 14; 49pp; English.
XX
XX The present sequence is the protein sequence of Streptococcus sobrinus
CC glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from S. mutants glucan binding protein-B (GbpB)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Diepitopic or
CC multiepitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX
SQ Sequence 1590 AA;

Query Match 100.0%; Score 85; DB 7; Length 1590;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDHL 14
Db |||||
337 QWNGESEKPYDDHL 350

RESULT 12
ADX37380
ID ADX37280 standard; protein; 1590 AA.
XX
AC ADX37280;
XX
XX 21-APR-2005 (first entry)
DT
XX
DE Streptococcus sobrinus glucan binding protein B.
XX
XX immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus sobrinus.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.

```

PR 12-APR-1999; 99US-00290049.  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 PR 07-MAR-2003; 2003US-00383930.  
 XX  
 PA (SMIT/) SMITH D J.  
 PA (TAUB/) TAUBMAN M A.  
 XX  
 XX Smith DJ, Taubman MA;  
 XX WPI; 2005-151644/16.  
 XX  
 XX New composition comprising a fragment of a glucan binding protein-B  
 PT (GbpB) that binds to MHC class II protein, and a biocompatible  
 PT microparticle, useful for producing an antibody (claimed) for immunizing  
 PT mammals against dental caries.  
 XX  
 PS Claim 7; SEQ ID NO 37; 73pp; English.  
 XX  
 CC The invention relates to a composition comprising a fragment of a glucan  
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II  
 CC protein. The composition is useful for producing an antibody for  
 CC immunizing mammals against dental caries. This sequence corresponds to a  
 CC Streptococcus sobrinus GbpB protein of the invention.  
 XX  
 SQ Sequence 1590 AA;  
 Query Match 100.0%; Score 85; DB 9; Length 1590;  
 Best Local Similarity 100.0%; Pred. No. 0.00016;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QWNGESEKPYDDHL 14  
 DB 337 QWNGESEKPYDDHL 350  
 |||||  
 RESULT 14  
 AAY43354  
 ID AAY43354 standard; peptide; 15 AA.  
 XX  
 XX AAY43354;  
 XX AC  
 XX DT 26-JAN-2000 (first entry)  
 XX DE GTF antigenic epitope.  
 XX  
 KW GTF; antigenic epitope; glucosyltransferase; immunogenic epitope; teeth;  
 KW immunogenic composition; streptococcal glucan; antibody response;  
 KW cariogenic bacteria; mutans streptococci; colonisation; caries;  
 KW dental caries; immunisation; therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9552548-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 XX 09-APR-1999; 99WO-US007828.  
 PF  
 XX 10-APR-1998; 98US-0081315P.  
 PR  
 XX (LEES/) LEES A.  
 PA (TAUB/) TAUBMAN M A.  
 PA (SMIT/) SMITH D J.  
 XX  
 PI Lees A, Taubman MA, Smith DJ;  
 XX WPI; 1999-620289/53.  
 DR  
 XX Immunogenic compositions for control of dental caries, based on  
 PT Streptococcus mutans components, particularly for vaccination of infants.  
 PS Disclosure; Page 17; 54pp; English.  
 XX  
 CC This sequence represents a glucosyltransferase (GTF) immunogenic epitope,  
 CC and can be used in the immunogenic composition of the invention. The  
 CC immunogenic composition (A) comprises at least one streptococcal glucan  
 CC (SG) and at least one component (I) covalently coupled, (in)directly, to  
 CC SG. (A) elicits antibodies (Ab) to both SG and (I). (A), and related  
 CC compositions, are used to induce an antibody response to cariogenic  
 CC bacteria, collectively called "mutans streptococci" (i.e. any of  
 CC Streptococcus mutans, S. rattus, S. cricetus or S. sobrinus), especially  
 CC to prevent colonisation, to reduce numbers of bacteria or to reduce  
 CC incidence of caries, in humans or other animals. The composition can  
 CC therefore be used for the treatment or prevention of caries (particularly  
 CC dental caries). Abs raised against (A) can be used: (a) to detect SG, and  
 CC (b) for passive immunisation, e.g. applied directly to the surface of  
 CC teeth. SG is a poorly immunogenic type 2 T-cell independent antigen but  
 CC when coupled to (I) it produces a conjugate which induces significant  
 CC immunity to both components, resulting in immunological memory and long-  
 CC lasting antibody production against SG, in adults and children  
 XX  
 SQ Sequence 15 AA;  
 Query Match 100.0%; Score 85; DB 9; Length 1590;  
 Best Local Similarity 100.0%; Pred. No. 0.00016;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QWNGESEKPYDDHL 14  
 DB 337 QWNGESEKPYDDHL 350  
 |||||  
 RESULT 14  
 AAR32925  
 ID AAR32925 standard; protein; 1592 AA.  
 XX  
 AC AAR32925;  
 XX  
 DT 28-JUN-1993 (first entry)  
 DE  
 DE Glucosyltransferase I.  
 XX  
 XX GT-1; Streptococcus; dental; caries.  
 KW  
 OS Streptococcus sobrinus.  
 XX  
 XX JP05023188-A.  
 PN  
 PD 02-FEB-1993.  
 XX  
 XX 25-JUL-1991; 91JP-00186592.  
 PF  
 XX 25-JUL-1991; 91JP-00186592.  
 PR  
 XX (KATO/) KATO K.  
 PA (FUKU/) FUKUI I.  
 XX  
 XX WPI; 1993-079449/10.  
 DR  
 DR N-PSDB; AAQ37760.  
 XX  
 XX DNA sequence glucosyltransferase-I - comprises Streptococcus sobrinus  
 PT DNA sequence with at least one nucleotide added or deleted.  
 PT  
 PS Claim 13; Page 15; 29pp; Japanese.  
 XX  
 CC The DNA sequence from Streptococcus sobrinus strain 6715 encodes  
 CC Glucosyltransferase-I (and mutants). The DNA was obcd. by treating S.  
 CC sobrinus 6715 with mutanolysin, extracting the chromosomal DNA, partially

CC digesting with Sau3AI and fractionating on agarose gel. The 3-5 kbp  
 CC fragment was ligated into pUC18 and E. coli JM109 transformed with it. A  
 CC GT-1 expressing clone was isolated and sequenced. The clone may be used  
 CC in the development of a drug for dental caries  
 XX  
 SQ Sequence 1592 AA;  
 Query Match 100.0%; Score 85; DB 2; Length 1592;  
 Best Local Similarity 100.0%; Pred. No. 0.00016;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 QWNGESEKPYDDHL 14  
 DB 337 QWNGESEKPYDDHL 350  
 |||||  
 RESULT 14  
 AAY43354  
 ID AAY43354 standard; peptide; 15 AA.  
 XX  
 XX AAY43354;  
 XX AC  
 XX DT 26-JAN-2000 (first entry)  
 XX DE GTF antigenic epitope.  
 XX  
 KW GTF; antigenic epitope; glucosyltransferase; immunogenic epitope; teeth;  
 KW immunogenic composition; streptococcal glucan; antibody response;  
 KW cariogenic bacteria; mutans streptococci; colonisation; caries;  
 KW dental caries; immunisation; therapy.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9552548-A2.  
 XX  
 PD 21-OCT-1999.  
 XX  
 XX 09-APR-1999; 99WO-US007828.  
 PF  
 XX 10-APR-1998; 98US-0081315P.  
 PR  
 XX (LEES/) LEES A.  
 PA (TAUB/) TAUBMAN M A.  
 PA (SMIT/) SMITH D J.  
 XX  
 PI Lees A, Taubman MA, Smith DJ;  
 XX WPI; 1999-620289/53.  
 DR  
 XX Immunogenic compositions for control of dental caries, based on  
 PT Streptococcus mutans components, particularly for vaccination of infants.  
 PS Disclosure; Page 17; 54pp; English.  
 XX  
 CC This sequence represents a glucosyltransferase (GTF) immunogenic epitope,  
 CC and can be used in the immunogenic composition of the invention. The  
 CC immunogenic composition (A) comprises at least one streptococcal glucan  
 CC (SG) and at least one component (I) covalently coupled, (in)directly, to  
 CC SG. (A) elicits antibodies (Ab) to both SG and (I). (A), and related  
 CC compositions, are used to induce an antibody response to cariogenic  
 CC bacteria, collectively called "mutans streptococci" (i.e. any of  
 CC Streptococcus mutans, S. rattus, S. cricetus or S. sobrinus), especially  
 CC to prevent colonisation, to reduce numbers of bacteria or to reduce  
 CC incidence of caries, in humans or other animals. The composition can  
 CC therefore be used for the treatment or prevention of caries (particularly  
 CC dental caries). Abs raised against (A) can be used: (a) to detect SG, and  
 CC (b) for passive immunisation, e.g. applied directly to the surface of  
 CC teeth. SG is a poorly immunogenic type 2 T-cell independent antigen but  
 CC when coupled to (I) it produces a conjugate which induces significant  
 CC immunity to both components, resulting in immunological memory and long-  
 CC lasting antibody production against SG, in adults and children  
 XX  
 SQ Sequence 15 AA;

Query Match 78.8%; Score 67; DB 2; Length 15;  
 Best Local Similarity 76.9%; Pred. No. 0.00074;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGESEKPYDDHL 14  
 || :||||:||||  
 Db 3 WNSDSEKPFDDHL 15

RESULT 15  
 ADV68556  
 ID ADV68556 standard; peptide; 15 AA.  
 XX  
 AC ADV68556;  
 XX  
 DT 24-FEB-2005 (first entry)  
 XX  
 DE S. mutans glucosyltransferase peptide GTF-C.  
 XX  
 KW Immunogenicity; vaccine; glucosyltransferase; GTF; tooth disease;  
 KW mouth disease; antibacterial; MAC peptide.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PN US6827936-B1.  
 XX  
 PD 07-DEC-2004.  
 XX  
 PF 12-APR-1999; 99US-00290049.  
 XX  
 PR 13-APR-1998; 98US-0081550P.  
 PR 08-JAN-1999; 99US-0115142P.  
 XX  
 PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX  
 DR WPI; 2005-009976/01.  
 XX  
 XX New immunogenic composition comprises a peptide corresponding to a  
 PT subunit of glucosyltransferase, useful as glucosyltransferase subunit  
 PT vaccine for preventing dental caries.  
 XX  
 PS Example 1; SEQ ID NO 6; 18pp; English.  
 XX  
 CC The present invention relates to a novel immunogenic composition which  
 CC comprises a peptide corresponding to a subunit of glucosyltransferase  
 CC (GTF) enzyme. The immunogenic composition is useful as a GTF subunit  
 CC vaccine for preventing dental caries. The present sequence is the  
 CC Streptococcus mutans glucosyltransferase peptide GTF-C.  
 XX  
 SQ Sequence 15 AA;

Query Match 78.8%; Score 67; DB 9; Length 15;  
 Best Local Similarity 76.9%; Pred. No. 0.00074;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGESEKPYDDHL 14  
 || :||||:||||  
 Db 3 WNSDSEKPFDDHL 15

RESULT 16  
 ABB98614  
 ID ABB98614 standard; peptide; 17 AA.  
 XX  
 AC ABB98614;  
 XX  
 XX 29-AUG-2003 (revised)  
 DT 14-JAN-2003 (first entry)  
 XX  
 DE Dextrane-saccharase, DSR-E, catalytic domain conserved peptide GTFB #1.

XX  
 KW Dextran saccharase; enzyme; cytostatic; dermatological; antiseborrheic;  
 KW DSR-E; glycosyl transferase; dextran; prebiotic; pharmaceutical;  
 KW microflora regulation; intestinal transit; mineral assimilation;  
 KW colon cancer; acne; dandruff; body odour.  
 XX  
 OS Leuconostoc mesenteroides; NRRL B-1299.  
 XX  
 PN FR2822163-A1.  
 XX  
 PD 20-SEP-2002.  
 XX  
 PF 19-DEC-2001; 2001FR-00016495.  
 XX  
 PR 16-MAR-2001; 2001FR-00003631.  
 XX  
 PA (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Bozonnet SAM, Renaud SMMC, Willemot RML, Monsan PEF;  
 XX  
 DR WPI; 2002-715213/78.  
 XX  
 PT New glycosyl transferase enzymes, containing glucan bonding and catalytic  
 PT domains and producing alpha-(1-2) branched dextrans, useful in probiotic,  
 PT pharmaceutical or cosmetic compositions.  
 XX  
 PS Example 2; Fig 5; 82pp; French.  
 XX  
 CC The present invention relates to a novel dextran saccharase, DSR-E  
 CC (ABB98574), from Leuconostoc mesenteroides NRRL B-1299. The saccharase  
 CC has glycosyl transferase activity suitable for producing dextrans having  
 CC alpha(1-2) branches from sucrose, p-nitrophenyl-alpha-D-glucopyranoside,  
 CC alpha-fluoroglucose, alpha-D-glucopyranoside-alpha-D-sorbiopyranoside or  
 CC alpha-D-galactopyranosyl-sucrose. The dextran saccharase is useful in  
 CC prebiotic, pharmaceutical or cosmetic compositions. The dextrans and  
 CC related compounds having alpha(1-2) bonds, produced using DSR-E, may be  
 CC involved in signalling/cellular recognition processes in vivo  
 CC (specifically in regulation of microflora in the intestines or on the  
 CC skin); and are potentially useful for improving intestinal transit,  
 CC increasing assimilation of minerals (e.g. calcium and/or magnesium),  
 CC preventing cancer of the colon and combating skin problems such as acne,  
 CC dandruff and body odour. The present sequence is a conserved peptide  
 CC sequence from the catalytic domain of the DSR-E of the invention.  
 CC (Updated on 29-AUG-2003 to standardise OS field)  
 XX  
 XX Sequence 17 AA;  
 SQ

Query Match 78.8%; Score 67; DB 5; Length 17;  
 Best Local Similarity 76.9%; Pred. No. 0.00085;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGESEKPYDDHL 14  
 || :||||:||||  
 Db 3 WNSDSEKPFDDHL 15

RESULT 17  
 AAR12630  
 ID AAR12630 standard; protein; 126 AA.  
 XX  
 AC AAR12630;  
 XX  
 DT 05-SEP-1991 (first entry)  
 XX  
 DE GtfB.1/CTB chimeric peptide antigenic for glucosyltransferase B protein.  
 XX  
 KW Dental carries; glucosyltransferase B; Cholera toxin; CTB.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9107979-A.  
 XX  
 PD 13-JUN-1991.



XX 29-NOV-1989; 89US-00442783.  
 XX  
 XX 29-NOV-1989; 89US-00442783.  
 XX  
 XX (INNO-) CENT INNOVATIVE TEC.  
 XX  
 XX Dertzbaugh MT, Macrina F;  
 XX  
 XX WPI; 1991-192968/26.  
 XX  
 XX N-PSDB; AAQ12262.  
 XX  
 XX New chimeric proteins for protection against dental caries - comprise the  
 PT epitope region of an antigen fused to the N-terminal of portion of the B  
 PT sub-unit of cholera toxin.  
 XX  
 XX Disclosure; Fig 11B; 66pp; English.  
 XX  
 XX The chimeric peptide is derived from the cholera toxin fused to the  
 CC glucosyltransferase B (GtfB) protein, it is non-toxic and gives rise to a  
 CC sustained antibody response. The product is antigenic for the GtfB  
 CC protein, and helps to prevent dental caries in the immunised host  
 XX  
 XX Sequence 126 AA;  
 SQ  
 Query Match 78.8%; Score 67; DB 2; Length 126;  
 Best Local Similarity 76.9%; Pred. No. 0.008;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 WNGSEKPYDDHL 14  
 DB || :|||:||||  
 6 WNSDSEKPFDDHL 18

RESULT 18  
 AAU79285  
 ID AAU79285 standard; protein; 1017 AA.  
 XX  
 XX AAU79285;  
 AC  
 DT 13-AUG-2002 (first entry)  
 XX  
 XX Streptococcus mutans monoclonal antibody-related protein #2.  
 DE  
 XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;  
 KW glucosyl transferase-B; immunotherapy.  
 XX  
 XX Streptococcus mutans.  
 OS  
 XX JP2002114709-A.  
 PN  
 XX 16-APR-2002.  
 PD  
 XX 04-OCT-2000; 2000JP-00304889.  
 PF  
 XX 04-OCT-2000; 2000JP-00304889.  
 PR  
 XX (UYN1-) UNIV NIPPON.  
 PA  
 XX WPI; 2002-448985/48.  
 DR  
 XX Anti-carries agent composed of a monoclonal antibody against an inhibitory  
 PT enzyme against water insoluble glucan synthetase of glucosyl transferase-  
 PT B (GTF-B) of Streptococcus mutans.  
 PT  
 XX Claim 4; Page 17-19; 28pp; Japanese.  
 PS  
 XX The invention relates to a monoclonal antibody against dental caries and  
 CC an anti-carries agent composed of a monoclonal antibody produced by  
 CC Streptococcus mutans, particularly mouse-hybridoma MHPi26 (FERM P-17566)  
 CC or mouse-hybridoma MHPi36 (FERM P-17567), against an enzyme having  
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
 CC transferase-B. The monoclonal antibody specifically inhibits water

CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl  
 CC transferase-B and is used in the immunotherapy of dental caries. This  
 CC sequence represents a Streptococcus mutans monoclonal antibody-related  
 XX protein  
 XX Sequence 1017 AA;  
 SQ  
 Query Match 78.8%; Score 67; DB 5; Length 1017;  
 Best Local Similarity 76.9%; Pred. No. 0.083;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 WNGSEKPYDDHL 14  
 DB || :|||:||||  
 310 WNSDSEKPFDDHL 322

RESULT 19  
 AAU98028  
 ID AAU98028 standard; protein; 1375 AA.  
 XX  
 XX AAU98028;  
 AC  
 XX 27-AUG-2002 (first entry)  
 DT  
 XX S. mutans glucosyltransferase GTFC.  
 DE  
 XX Glucosyltransferase; GTFC; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture.  
 XX  
 XX Streptococcus mutans.  
 OS  
 XX US2002031826-A1.  
 PN  
 XX 14-MAR-2002.  
 PD  
 XX 19-DEC-2000; 2000US-00740274.  
 PF  
 XX 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 18-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 XX (NICH/) NICHOLS S E.  
 PA  
 XX Nichols SE;  
 PI  
 XX WPI; 2002-414332/44.  
 DR N-PSDB; ABK52939.  
 DR  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 PT  
 XX Disclosure; Page 30-33; 44pp; English.  
 PS  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,

thermoplastic molecule or their combinations or glucan and starch where the glucan is produced in the amyloplast and/or vacuole or a maize line deficient in starch biosynthesis, transformed with a gene encoding a glucosyltransferase B or D enzyme, wild-type or mutant and a paper comprising the glucan (paper sizing/coating agent). The vector is useful for producing a glucan in a plant. The method comprises transforming a plant cell with the vector, growing the plant cell under plant growing conditions to produce a regenerated plant and inducing expression of the polynucleotide for a time sufficient to produce the glucan in the regenerated plant, where the vector contains a transit sequence from ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and chlorophyll AB binding protein to produce a transgenic plant, and Glucan is produced in the amyloplast of potato or the vacuole of sugar beet. Glucans are useful as substitutes for and additions to modified starch and latexes in paper manufacture. Unlike prior art techniques, which require input materials that produce chemical effluents, paper manufacture utilising the glucan produced by GTF, which utilises biologically produced input materials, is more cost-effective and environmentally friendly. Moreover, glucans also exhibit thermoplastic properties and impart gloss to the paper during coating step. The present sequence represents GTFc

XX  
SQ Sequence 1375 AA;

Query Match 78.8%; Score 67; DB 5; Length 1375;  
Best Local Similarity 76.9%; Pred. No. 0.12;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGESEKPYDDHL 14  
||:||||:||||  
Db 370 WNSDSEKPFDDHL 382

RESULT 20  
AAU79288  
ID AAU79288 standard; protein; 1375 AA.

AC AAU79288;

DT 13-AUG-2002 (first entry)

XX Streptococcus mutans monoclonal antibody-related protein #5.

KW Antibody; dental caries; water insoluble glucan synthetase; anti-carries;  
KW glucosyl transferase-B; immunotherapy.

OS Streptococcus mutans.

XX JF2002114709-A.

PD 16-APR-2002.

PF 04-OCT-2000; 2000JP-00304889.

PR 04-OCT-2000; 2000JP-00304889.

PA (UYNI-) UNIV NIPPON.

XX WPI; 2002-448885/48.

XX Anti-carries agent composed of a monoclonal antibody against an inhibitory  
PT enzyme against water insoluble glucan synthetase of glucosyl transferase-  
PT B (GTF-B) of Streptococcus mutans.

PS Disclosure; Page 22-25; 28pp; Japanese.

XX The invention relates to a monoclonal antibody against dental caries and  
CC an anti-carries agent composed of a monoclonal antibody produced by  
CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)  
CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having  
CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
CC transferase-B. The monoclonal antibody specifically inhibits water  
CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl

CC transferase-B and is used in the immunotherapy of dental caries. This  
CC sequence represents a Streptococcus mutans monoclonal antibody-related  
CC protein  
XX  
SQ Sequence 1375 AA;

Query Match 78.8%; Score 67; DB 5; Length 1375;  
Best Local Similarity 76.9%; Pred. No. 0.12;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGESEKPYDDHL 14  
||:||||:||||  
Db 370 WNSDSEKPFDDHL 382

RESULT 21

ID ADD93655 standard; protein; 1375 AA.

XX ADD93655;

DT 29-JAN-2004 (first entry)

XX Streptococcus mutans glucosyltransferase-C.

XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

OS Streptococcus mutans.

XX WO2003075845-A2.

PD 18-SEP-2003.

PF 07-MAR-2003; 2003WO-US006962.

XX 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman MA;

DR WPI; 2003-845091/78.

XX Composition useful as vaccines for dental caries comprises a fragment of  
PT a glucan binding protein-B binding to a major histocompatibility complex  
PT class II protein.

XX Claim 16; Page 13; 49pp; English.

XX The present sequence is the protein sequence of Streptococcus mutans  
CC glucosyltransferase-C (GTF-C). Peptide fragments of GTF-C, especially  
CC from the catalytic domain of the polypeptide, can be used in immunogenic  
CC compositions and subunit vaccines for dental caries. These compositions  
CC comprise a major histocompatibility complex (MHC) class II protein-  
CC binding peptide from S. mutans glucan binding protein-B (GbpB)  
CC covalently linked with a peptide fragment of a streptococcal  
CC glucosyltransferase. The compositions are used in a claimed method of  
CC eliciting production of an antibody in a mammal. Dieptopic or  
CC multicentric polypeptides can be prepared synthetically or by  
CC recombinant DNA technology. Antibodies raised against MHC class II  
CC binding fragments of GbpB can be used in passive immunisation.

XX Sequence 1375 AA;

Query Match 78.8%; Score 67; DB 7; Length 1375;  
Best Local Similarity 76.9%; Pred. No. 0.12;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGESEKPYDDHL 14  
||:||||:||||  
Db 370 WNSDSEKPFDDHL 382

## RESULT 22

ADX37278  
 XX ADX37278 standard; protein; 1375 AA.  
 AC  
 XX  
 XX ADX37278;  
 DT 21-APR-2005 (first entry)  
 XX  
 DE Streptococcus mutant glucan binding protein B variant #7.  
 XX  
 XX immunogenicity; immune stimulation; glucan binding protein-B;  
 KW microparticle; major histocompatibility complex; tooth disease.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PN US2005031633-A1.  
 XX  
 PD 10-FEB-2005.  
 XX  
 PF 09-MAR-2004; 2004US-00797821.  
 XX  
 PR 13-APR-1998; 98US-0081550P.  
 PR 08-JAN-1999; 99US-0115142P.  
 PR 12-APR-1999; 99US-00290049.  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 PR 07-MAR-2003; 2003US-00383930.  
 XX  
 PA (SMIT/) SMITH D J.  
 PA (TAUB/) TAUBMAN M A.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX  
 XX WPI; 2005-151644/16.  
 DR  
 XX  
 XX New composition comprising a fragment of a glucan binding protein-B  
 PT (GbpB) that binds to MHC class II protein, and a biocompatible  
 PT microparticle, useful for producing an antibody (claimed) for immunizing  
 PT mammals against dental caries.  
 XX  
 XX Claim 7; SEQ ID NO 35; 73pp; English.  
 PS  
 XX The invention relates to a composition comprising a fragment of a glucan  
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II  
 CC protein. The composition is useful for producing an antibody for  
 CC immunizing mammals against dental caries. This sequence corresponds to a  
 CC Streptococcus mutans GbpB protein of the invention.  
 XX  
 SQ Sequence 1375 AA;  
 CC  
 Query Match 78.8%; Score 67; DB 9; Length 1375;  
 Best Local Similarity 76.9%; Pred. No. 0.12;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 WNGSEKPYDDHL 14  
 || :|||:||||  
 Db 370 WNSDSEKPPDDHL 382  
 RESULT 23  
 AAU98031  
 ID AAU98031 standard; protein; 1475 AA.  
 XX  
 AC AAU98031;  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE S. mutans glucosyltransferase GTFB mutant D457N.  
 XX  
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture; mutant; mutein.  
 XX  
 OS Streptococcus mutans.  
 OS Synthetic.  
 PH  
 FT Key Location/Qualifiers  
 FT Misc-difference 457 /note= "Wild-type Asp substituted by Asn"  
 XX  
 PN US2002031826-A1.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PF 19-DEC-2000; 2000US-00740274.  
 XX  
 PR 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 PA (NICH/) NICHOLS S E.  
 XX  
 XX Nichols SE;  
 PI  
 XX WPI; 2002-414332/44.  
 DR  
 XX  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 PS Claim 36; Page; 44pp; English.  
 XX  
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I48V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36

```

XX SQ Sequence 1475 AA;
Query Match 78.8%; Score 67; DB 5; Length 1475;
Best Local Similarity 76.9%; Pred. No. 0.13;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGESEKPYDDHL 14
DB 344 WNSDSEKPFDDHL 356
||:||||:||||
||:||||:||||

RESULT 24
AAU98036
ID AAU98036 standard; protein; 1475 AA.
AC AAU98036;
XX
DT 27-AUG-2002 (first entry)
DE S. mutans glucosyltransferase GTFB mutant D567T/D571K.
XX
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
XX
PN US2002031826-A1.
XX
PD 14-MAR-2002.
XX
PF 19-DEC-2000; 2000US-00740274.
XX
PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-0007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX
PA (NICH/) NICHOLS S E.
XX
PI Nichols SE;
XX
XX WPI; 2002-414332/44.
XX
PT Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
XX Claim 36; Page; 44pp; English.
XX
PS
PS
CC The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I448V, D457N, D567T,
CC K1014I, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014I,
CC I448V/D457N/D567T/D571K/K779Q/K1014I, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or

```

```

CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFB mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
CC claim 36
XX
SQ Sequence 1475 AA;

Query Match 78.8%; Score 67; DB 5; Length 1475;
Best Local Similarity 76.9%; Pred. No. 0.13;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGESEKPYDDHL 14
DB 344 WNSDSEKPFDDHL 356
||:||||:||||
||:||||:||||

RESULT 25
AAU98037
ID AAU98037 standard; protein; 1475 AA.
XX
AC AAU98037;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFB mutant D567T/D571K/K1014T.
XX
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"
XX
PN US2002031826-A1.
XX
PD 14-MAR-2002.
XX
PF 19-DEC-2000; 2000US-00740274.
XX
PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-0007999.

```

PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 PA (NICH/) NICHOLS S E.  
 XX Nichols SE;  
 XX WPI; 2002-414332/44.  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Claim 36; Page; 44pp; English.  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX  
 SQ Sequence 1475 AA;  
 Query Match 78.8%; Score 67; DB 5; Length 1475;  
 Best Local Similarity 76.9%; Pred. No. 0.13;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 WNGSEKPYDDHL 14  
 || :||||:||||  
 Db 344 WNSDSEKPFDDHL 356  
 RESULT 26  
 AAU98040  
 ID AAU98040 standard; protein; 1475 AA.  
 XX  
 AC AAU98040;  
 XX  
 DT 27-AUG-2002 (first entry)

XX S. mutans glucosyltransferase GTFB mutant K779Q.  
 DE Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 XX coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture; mutant; mutain.  
 XX Streptococcus mutans.  
 OS Synthetic.  
 OS  
 XX Key Location/Qualifiers  
 PH Misc-difference 779 /note= "Wild-type Lys substituted by Gln"  
 FT  
 FT  
 XX US2002031826-A1.  
 PN  
 XX 14-MAR-2002.  
 PD  
 XX 19-DEC-2000; 2000US-00740274.  
 PF  
 XX 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 18-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 PA  
 XX Nichols SE;  
 FI  
 XX WPI; 2002-414332/44.  
 DR  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Claim 36; Page; 44pp; English.  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX  
 SQ Sequence 1475 AA;  
 Query Match 78.8%; Score 67; DB 5; Length 1475;  
 Best Local Similarity 76.9%; Pred. No. 0.13;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 WNGSEKPYDDHL 14  
 || :||||:||||  
 Db 344 WNSDSEKPFDDHL 356  
 RESULT 26  
 AAU98040  
 ID AAU98040 standard; protein; 1475 AA.  
 XX  
 AC AAU98040;  
 XX  
 DT 27-AUG-2002 (first entry)

CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX

SQ Sequence 1475 AA;

Query Match 78.8%; Score 67; DB 5; Length 1475;

Best Local Similarity 76.9%; Pred. No. 0.13;

Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGESEKPYDDHL 14

Db 344 WNSDSEKPFDDHL 356

|| :||||:||||

RESULT 27

AAU98035

ID AAU98035 standard; protein; 1475 AA.

AC AAU98035;

XX 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFB mutant D457N/D571K.

XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;

KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX Streptococcus mutans.

OS Synthetic.

XX Key Location/Qualifiers

FH Misc-difference 457

FT /note= "Wild-type Asp substituted by Asn"

FT Misc-difference 571

FT /note= "Wild-type Asp substituted by Lys"

XX US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

XX 07-JUN-1995; 95US-00482711.

XX 07-JUN-1995; 95US-00485243.

XX 16-JAN-1998; 98US-00007999.

XX 16-JAN-1998; 98US-00008172.

XX 20-JAN-1998; 98US-00009620.

XX 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;

XX WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its

CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX

SQ Sequence 1475 AA;

Query Match 78.8%; Score 67; DB 5; Length 1475;

Best Local Similarity 76.9%; Pred. No. 0.13;

Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGESEKPYDDHL 14

Db 344 WNSDSEKPFDDHL 356

|| :||||:||||

RESULT 28

AAU98033

ID AAU98033 standard; protein; 1475 AA.

XX AAU98033;

XX 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFB mutant K1014T.

XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;

KW coating composition; glucan; starch; latex; thermoplastic molecule;

KW amyloplast; vacuole; paper manufacture; mutant; mutein.

XX Streptococcus mutans.

OS Synthetic.

XX Key Location/Qualifiers

FH Misc-difference 1014

FT /note= "Wild-type Lys substituted by Thr"

XX US2002031826-A1.

XX 14-MAR-2002.

XX 19-DEC-2000; 2000US-00740274.

XX 07-JUN-1995; 95US-00478704.

XX 07-JUN-1995; 95US-00482711.

XX 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 XX Nichols SE;  
 PI WPI; 2002-414332/44.  
 DR Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Claim 36; Page; 44pp; English.  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX SQ Sequence 1475 AA;  
 Query Match 78.8%; Score 67; DB 5; Length 1475;  
 Best Local Similarity 76.9%; Pred. No. 0.13;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 WNGSEKPYDDHL 14  
 ||:|||||  
 Db 344 WNSDSEKPFDDHL 356  
 RESULT 29  
 AAU98034  
 ID AAU98034 standard; protein; 1475 AA.  
 XX  
 AC AAU98034;  
 XX

DT 27-AUG-2002 (first entry)  
 XX S. mutans glucosyltransferase GTFB mutant D457N/D567T.  
 DE Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 XX coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture; mutant; mutain.  
 KW Streptococcus mutans.  
 XX Synthetic.  
 OS Key Location/Qualifiers  
 FH Misc-difference 457 /note= "Wild-type Asp substituted by Asn"  
 FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"  
 FT US2002031826-A1.  
 PN 14-MAR-2002.  
 XX 19-DEC-2000; 2000US-00740274.  
 XX 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 PA Nichols SE;  
 PI WPI; 2002-414332/44.  
 DR Glucosyltransferase B or D protein useful for producing a glucan useful  
 XX as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 PT Claim 36; Page; 44pp; English.  
 PS The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX SQ Sequence 1475 AA;  
 Query Match 78.8%; Score 67; DB 5; Length 1475;  
 Best Local Similarity 76.9%; Pred. No. 0.13;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 WNGSEKPYDDHL 14  
 ||:|||||  
 Db 344 WNSDSEKPFDDHL 356  
 RESULT 29  
 AAU98034  
 ID AAU98034 standard; protein; 1475 AA.  
 XX  
 AC AAU98034;  
 XX

CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 CC  
 XX Sequence 1475 AA;  
 SQ

Query Match 78.8%; Score 67; DB 5; Length 1475;  
 Best Local Similarity 76.9%; Pred. No. 0.13;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14  
 || :||||:||||  
 Db 344 WNSDSEKPFDDHL 356

RESULT 30  
 AAU98030  
 ID AAU98030 standard; protein; 1475 AA.  
 XX  
 AC AAU98030;  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE S. mutans glucosyltransferase GTFB mutant I448V.  
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.  
 XX  
 OS Streptococcus mutans.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 448 /note= "Wild-type Ile substituted by Val"  
 FT  
 XX  
 PN US2002031826-A1.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PF 19-DEC-2000; 2000US-00740274.  
 XX  
 PR 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 PA (NICH/) NICHOLS S E.  
 XX  
 XX Nichols SE;  
 PI  
 DR WPI; 2002-414332/44.  
 XX  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 PS Claim 36; Page; 44pp; English.  
 XX  
 CC The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the

CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 CC  
 XX Sequence 1475 AA;  
 SQ

Query Match 78.8%; Score 67; DB 5; Length 1475;  
 Best Local Similarity 76.9%; Pred. No. 0.13;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14  
 || :||||:||||  
 Db 344 WNSDSEKPFDDHL 356

RESULT 31  
 AAU98039  
 ID AAU98039 standard; protein; 1475 AA.  
 XX  
 AC AAU98039;  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE S. mutans glucosyltransferase GTFB mutant YYY169-171AAA.  
 KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture; mutant; mutein.  
 XX  
 OS Streptococcus mutans.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 169..171 /note= "Wild-type Tyr-Tyr substituted by Ala-Ala-Ala"  
 FT  
 XX  
 PN US2002031826-A1.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PF 19-DEC-2000; 2000US-00740274.  
 XX  
 PR 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.



PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 XX Nichols SE;  
 XX WPI; 2002-414332/44.  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Claim 36; Page; 44pp; English.  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX SQ Sequence 1475 AA;  
 Query Match 78.8%; Score 67; DB 5; Length 1475;  
 Best Local Similarity 76.9%; Pred. No. 0.13;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 WNGSEKPYDDHL 14  
 Db 344 WNSDSEKPFDDHL 356  
 RESULT 32  
 AAU98027  
 ID AAU98027 standard; protein; 1475 AA.  
 XX  
 AC AAU98027;

XX 27-AUG-2002 (first entry)  
 XX S. mutans glucosyltransferase GTFB.  
 XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture.  
 XX Streptococcus mutans.  
 XX US2002031826-A1.  
 XX 14-MAR-2002.  
 XX 19-DEC-2000; 2000US-00740274.  
 XX 07-JUN-1995; 95US-00478704.  
 XX 07-JUN-1995; 95US-00482711.  
 XX 07-JUN-1995; 95US-00485243.  
 XX 16-JAN-1998; 98US-00007999.  
 XX 16-JAN-1998; 98US-00008172.  
 XX 20-JAN-1998; 98US-00009620.  
 XX 11-DEC-1998; 98US-00210361.  
 XX (NICH/) NICHOLS S E.  
 XX Nichols SE;  
 XX WPI; 2002-414332/44.  
 XX N-PSDB; ABK52938.  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX Disclosure; Page 21-25; 44pp; English.  
 XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36  
 XX SQ Sequence 1475 AA;  
 Query Match 78.8%; Score 67; DB 5; Length 1475;  
 Best Local Similarity 76.9%; Pred. No. 0.13;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 WNGSEKPYDDHL 14  
 Db 344 WNSDSEKPFDDHL 356  
 RESULT 32  
 AAU98027  
 ID AAU98027 standard; protein; 1475 AA.  
 XX  
 AC AAU98027;

```

XX SQ Sequence 1475 AA;
Query Match 78.8%; Score 67; DB 5; Length 1475;
Best Local Similarity 76.9%; Pred. No. 0.13;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14
DB 344 WNSDSEKPFDDHL 356

RESULT 33
ID AAU98032 standard; protein; 1475 AA.
XX AC AAU98032;
XX DT 27-AUG-2002 (first entry)
XX DE S. mutans glucosyltransferase GTFB mutant D567T.
XX KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
XX KW coating composition; glucan; starch; latex; thermoplastic molecule;
XX KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX OS Streptococcus mutans.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
FT FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
FT FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
FT FT Misc-difference 779 /note= "Wild-type Lys substituted by Gln"
FT FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"
XX US2002031826-A1.
XX PD 14-MAR-2002.
XX PF 19-DEC-2000; 2000US-00740274.
XX

XX SQ Sequence 1475 AA;
Query Match 78.8%; Score 67; DB 5; Length 1475;
Best Local Similarity 76.9%; Pred. No. 0.13;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14
DB 344 WNSDSEKPFDDHL 356

RESULT 34
AAU98038
ID AAU98038 standard; protein; 1475 AA.
XX AC AAU98038;
XX DT 27-AUG-2002 (first entry)
XX DE S. mutans GTFB mutant I448V/D457N/D567T/D571K/K1014T.
XX KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
XX KW coating composition; glucan; starch; latex; thermoplastic molecule;
XX KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX OS Streptococcus mutans.
XX OS Synthetic.
XX FH Key Location/Qualifiers
FT Misc-difference 448 /note= "Wild-type Ile substituted by Val"
FT FT Misc-difference 457 /note= "Wild-type Asp substituted by Asn"
FT FT Misc-difference 567 /note= "Wild-type Asp substituted by Thr"
FT FT Misc-difference 571 /note= "Wild-type Asp substituted by Lys"
FT FT Misc-difference 779 /note= "Wild-type Lys substituted by Gln"
FT FT Misc-difference 1014 /note= "Wild-type Lys substituted by Thr"
XX US2002031826-A1.
XX PD 14-MAR-2002.
XX PF 19-DEC-2000; 2000US-00740274.
XX

```

```
PR 07-JUN-1995; 95US-00478704.
PR 07-JUN-1995; 95US-00482711.
PR 07-JUN-1995; 95US-00485243.
PR 16-JAN-1998; 98US-00007999.
PR 16-JAN-1998; 98US-00008172.
PR 20-JAN-1998; 98US-00009620.
PR 11-DEC-1998; 98US-00210361.
XX (NICH//) NICHOLS S E.
XX Nichols SE;
XX WPI; 2002-414332/44.
XX
XX Glucosyltransferase B or D protein useful for producing a glucan useful
PT as substitutes for and additions to modified starch and latexes in paper
PT manufacture, comprises mutations in specific positions.
XX
XX Claim 36; Page; 44pp; English.
XX
XX The invention an isolated protein comprising a glucosyltransferase (GTF)
CC B polypeptide having changes at position from I448V, D457N, D567T,
CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,
CC I448V/D457N/D567T/D571K/K1014T, Y169A/Y170A/Y171A, and K779Q or a
CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,
CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the
CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its
CC complementary polynucleotide, a ribonucleic acid sequence encoding the
CC GTF mutant, an expression cassette comprising the polynucleotide operably
CC linked to a promoter, a vector comprising the expression cassette, host
CC cell introduced with the vector, a transgenic plant comprising the
CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or
CC coating composition comprising a glucan produced in a plant transformed
CC with a gene encoding the mutant GTF, wild type or, starch, a latex,
CC thermoplastic molecule or their combinations or glucan and starch where
CC the glucan is produced in the amyloplast and/or vacuole or a maize line
CC deficient in starch biosynthesis, transformed with a gene encoding a
CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper
CC comprising the glucan (paper sizing/coating agent). The vector is useful
CC for producing a glucan in a plant. The method comprises transforming a
CC plant cell with the vector, growing the plant cell under plant growing
CC conditions to produce a regenerated plant and inducing expression of the
CC polynucleotide for a time sufficient to produce the glucan in the
CC regenerated plant, where the vector contains a transit sequence from
CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and
CC chlorophyll AB binding protein to produce a transgenic plant, and glucan
CC is produced in the amyloplast of potato or the vacuole of sugar beet.
CC Glucans are useful as substitutes for and additions to modified starch
CC and latexes in paper manufacture. Unlike prior art techniques, which
CC require input materials that produce chemical effluents, paper
CC manufacture utilising the glucan produced by GTF, which utilises
CC biologically produced input materials, is more cost-effective and
CC environmentally friendly. Moreover, glucans also exhibit thermoplastic
CC properties and impart gloss to the paper during coating step. The present
CC sequence represents a GTFB mutant of the invention. Note: The present
CC sequence is not shown in the specification but was created by the indexer
CC using the GTFB sequence appearing as AAU98027 and the information in
CC claim 36
XX
XX Sequence 1475 AA;
XX
XX Query Match 78.8%; Score 67; DB 5; Length 1475;
XX Best Local Similarity 76.9%; Pred. No. 0.13;
XX Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
XX
XX Qy 2 WNGSEKPYDDHL 14
XX || :|||:|
XX Db 344 WNSDSEKPFDDHL 356
XX
XX RESULT 35
XX ADD93654
XX ID ADD93654 standard; protein; 1475 AA.
XX
```

---

```
XX ADD93654;
XX
XX 29-JAN-2004 (first entry)
XX
XX Streptococcus mutans glucosyltransferase-B.
XX
XX Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
XX Streptococcus mutans.
XX
XX WO2003075845-A2.
XX
XX 18-SEP-2003.
XX
XX 07-MAR-2003; 2003WO-US006962.
XX
XX 07-MAR-2002; 2002US-0363209P.
XX
XX 08-AUG-2002; 2002US-0402483P.
XX
XX (FORS-) FORSYTH INST.
XX
XX Smith DJ, Taubman MA;
XX
XX WPI; 2003-845091/78.
XX
XX Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
XX Claim 16; Page 12-13; 49pp; English.
XX
XX The present sequence is the protein sequence of Streptococcus mutans
CC glucosyltransferase-B (GTF-B). Peptide fragments of GTF-B, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from S. mutans glucan binding protein-B (GbpB)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Diepitopic or
CC multipitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX
XX Sequence 1475 AA;
XX
XX Query Match 78.8%; Score 67; DB 7; Length 1475;
XX Best Local Similarity 76.9%; Pred. No. 0.13;
XX Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
XX
XX Qy 2 WNGSEKPYDDHL 14
XX || :|||:|
XX Db 344 WNSDSEKPFDDHL 356
XX
XX RESULT 36
XX ADX37277
XX ID ADX37277 standard; protein; 1475 AA.
XX
XX AC ADX37277;
XX
XX 21-APR-2005 (first entry)
XX
XX Streptococcus mutant glucan binding protein B variant #6.
XX
XX immunogenicity; immune stimulation; glucan binding protein-B;
XX microparticle; major histocompatibility complex; tooth disease.
XX
XX Streptococcus mutans.
XX
XX US2005031633-A1.
XX
```

```
PD 10-FEB-2005.
XX
XX 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
XX Smith DJ, Taubman MA;
XX
XX WPI; 2005-151644/16.
DR
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX Claim 7; SEQ ID NO 34; 73pp; English.
PS
XX
XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
XX Sequence 1475 AA;
SQ
Query Match 78.8%; Score 67; DB 9; Length 1475;
Best Local Similarity 76.9%; Pred. No. 0.13;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 WNGSEKPYDDHL 14
DB 344 WNSDSEKPFDDHL 356
RESULT 37
AAU79284
ID AAU79284 standard; protein; 1476 AA.
XX
XX AAU79284;
AC
XX
XX 13-AUG-2002 (first entry)
DT
XX
XX Streptococcus mutans monoclonal antibody-related protein #1.
DE
XX
XX Antibody; dental caries; water insoluble glucan synthetase; anti-carries;
KW glucosyl transferase-B; immunotherapy.
KW
XX
XX Streptococcus mutans.
OS
XX
XX JP2002114709-A.
PN
XX
XX 16-APR-2002.
PD
XX
XX 04-OCT-2000; 2000JP-00304889.
PF
XX
XX 04-OCT-2000; 2000JP-00304889.
PR
XX
XX (UYNI-) UNIV NIPPON.
PA
XX
XX WPI; 2002-448885/48.
DR
XX
XX Anti-carries agent composed of a monoclonal antibody against an inhibitory
PT enzyme against water insoluble glucan synthetase of glucosyl transferase-
PT B (GTF-B) of Streptococcus mutans.
us-10-797-821-28.rag
XX
XX Claim 3; Page 13-16; 28pp; Japanese.
XX
XX The invention relates to a monoclonal antibody against dental caries and
CC an anti-carries agent composed of a monoclonal antibody produced by
CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)
CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having
CC inhibitive activity against water insoluble glucan synthetase of glucosyl
CC transferase-B. The monoclonal antibody specifically inhibits water
CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl
CC transferase-B and is used in the immunotherapy of dental caries. This
CC sequence represents a Streptococcus mutans monoclonal antibody-related
CC protein
XX
XX Sequence 1476 AA;
SQ
Query Match 78.8%; Score 67; DB 5; Length 1476;
Best Local Similarity 76.9%; Pred. No. 0.13;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 WNGSEKPYDDHL 14
DB 344 WNSDSEKPFDDHL 356
RESULT 38
ADV68555
ID ADV68555 standard; peptide; 15 AA.
XX
XX ADV68555;
AC
XX
XX 24-FEB-2005 (first entry)
DT
XX
XX S. mutans glucosyltransferase peptide GTF-B.
DE
XX
XX Immunogenicity; vaccine; glucosyltransferase; GTF; tooth disease;
KW mouth disease; antibacterial; MAC peptide.
KW
XX
XX Streptococcus mutans.
OS
XX
XX US6827936-B1.
PN
XX
XX 07-DEC-2004.
PD
XX
XX 12-APR-1999; 99US-00290049.
PF
XX
XX 13-APR-1998; 98US-0081550P.
PR
XX
XX 08-JAN-1999; 99US-0115142P.
PR
XX
XX (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
PA
XX
XX Smith DJ, Taubman MA;
PI
XX
XX WPI; 2005-009976/01.
DR
XX
XX New immunogenic composition comprises a peptide corresponding to a
PT subunit of glucosyltransferase, useful as glucosyltransferase subunit
PT vaccine for preventing dental caries.
PT
XX
XX Example 1; SEQ ID NO 5; 18pp; English.
PS
XX
XX The present invention relates to a novel immunogenic composition which
CC comprises a peptide corresponding to a subunit of glucosyltransferase
CC (GTF) enzyme. The immunogenic composition is useful as a GTF subunit
CC vaccine for preventing dental caries. The present sequence is the
CC Streptococcus mutans GTF-B control peptide, MAC.
XX
XX Sequence 15 AA;
SQ
Query Match 75.3%; Score 64; DB 9; Length 15;
Best Local Similarity 69.2%; Pred. No. 0.0023;
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```

Qy      2 WNGESEKPYDDHL 14
Db      3 WNSDSERPFDDHL 15

RESULT 39
AB98619
ID      ABB98619 standard; peptide; 17 AA.
XX      AC
XX      ABB98619;
DT      29-AUG-2003 (revised)
DT      14-JAN-2003 (first entry)
XX      XX
XX      Dextrane-saccharase, DSR-E, catalytic domain conserved peptide dsrB #1.
XX      XX
XX      Dextran saccharase; enzyme; cytostatic; dermatological; antiseborrheic;
XX      DSR-E; glycosyl transferase; dextran; prebiotic; pharmaceutical;
XX      microflora regulation; intestinal transit; mineral assimilation;
XX      colon cancer; acne; dandruff; body odour.
XX      XX
XX      Leuconostoc mesenteroides; NRRL B-1299.
XX      XX
XX      FR2822163-A1.
XX      PD
XX      20-SEP-2002.
XX      XX
XX      19-DEC-2001; 2001FR-00016495.
XX      XX
XX      16-MAR-2001; 2001FR-00003631.
XX      XX
XX      (CNRS ) CNRS CENT NAT RECH SCI.
XX      XX
XX      Bozonnet SAM, Remaud SMMC, Willemot RML, Monsan PEF;
XX      XX
XX      WPI; 2002-715213/78.
XX      XX
XX      New glycosyl transferase enzymes, containing glucan bonding and catalytic
XX      domains and producing alpha-(1-2) branched dextrans, useful in probiotic,
XX      pharmaceutical or cosmetic compositions.
XX      XX
XX      Example 2; Fig 5; 82pp; French.
XX      XX
XX      The present invention relates to a novel dextran saccharase, DSR-E
XX      (AB98574), from Leuconostoc mesenteroides NRRL B-1299. The saccharase
XX      has glycosyl transferase activity suitable for producing dextrans having
XX      alpha(1-2) branches from sucrose. p-nitrophenyl-alpha-D-glucopyranoside,
XX      alpha-fluoroglucose, alpha-D-glucopyranoside-alpha-D-sorbypyranoside or
XX      alpha-D-galactopyranosyl-sucrose. The dextran saccharase is useful in
XX      prebiotic, pharmaceutical or cosmetic compositions. The dextrans and
XX      related compounds having alpha(1-2) bonds, produced using DSR-E, may be
XX      involved in signalling/cellular recognition processes in vivo
XX      (specifically in regulation of microflora in the intestines or on the
XX      skin); and are potentially useful for improving intestinal transit,
XX      increasing assimilation of minerals (e.g. calcium and/or magnesium),
XX      preventing cancer of the colon and combating skin problems such as acne,
XX      dandruff and body odour. The present sequence is a conserved peptide
XX      sequence from the catalytic domain of the DSR-E of the invention.
XX      (Updated on 29-AUG-2003 to standardise OS field)
XX      XX
XX      Sequence 17 AA;

Query Match      60.0%; Score 51; DB 5; Length 17;
Best Local Similarity 64.3%; Pred. No. 0.35;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy      1 QWNGESEKPYDDHL 14
Db      2 QWNMSSEDPKNDHL 15

RESULT 40
ADY72733
ID      ADY72733
XX      AC
XX      19-MAY-2005 (first entry)
XX      DT
XX      Mutant dextranucrase #2.
XX      KW
XX      muten; dextranucrase; enzyme.
XX      OS
XX      Unidentified.
XX      PN
XX      KR2004092249-A.
XX      PD
XX      03-NOV-2004.
XX      PF
XX      25-APR-2003; 2003KR-00026494.
XX      PR
XX      25-APR-2003; 2003KR-00026494.
XX      PA
XX      (KIMD/) KIM D M.
XX      PI
XX      Kang HG, Kim CY, Kim DM, Seo ES;
XX      DR
XX      WPI; 2005-178323/19.
XX      DR
XX      N-PSDB; ADY72731.
XX      XX
XX      Method for preparing mutant gene by using, as mutagen, radiation selected
XX      from UVA(ultra violet a), UVB(ultra violet b), UVC(ultra violet c), x-
XX      ray, VUV(vacuum ultra violet) and ultrasoft x-ray.
XX      XX
XX      Disclosure; Fig 3; 20pp; Korean.
XX      XX
XX      The invention relates to a method for preparing a mutant gene by
XX      radiation is provided, to mutate a specific gene selectively by direct
XX      radiation. Therefore, it can be useful for improving genes of industrial
XX      enzymes and developing enzymes having novel properties. The present
XX      sequence represents a dextranucrase mutant DNA.
XX      XX
XX      Sequence 1477 AA;

Query Match      60.0%; Score 51; DB 9; Length 1477;
Best Local Similarity 64.3%; Pred. No. 52;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy      1 QWNGESEKPYDDHL 14
Db      397 QWNMSSEDPKNDHL 410

RESULT 41
ADY72732
ID      ADY72732 standard; protein; 1477 AA.
XX      AC
XX      ADY72732;
XX      DT
XX      19-MAY-2005 (first entry)
XX      DE
XX      Mutant dextranucrase #1.
XX      KW
XX      muten; dextranucrase; enzyme.
XX      OS
XX      Unidentified.
XX      PN
XX      KR2004092249-A.
XX      PD
XX      03-NOV-2004.
XX      PF
XX      25-APR-2003; 2003KR-00026494.
XX      PR
XX      25-APR-2003; 2003KR-00026494.
XX      PA
XX      (KIMD/) KIM D M.

```

XX Kang HG, Kim CY, Kim DM, Seo ES;  
XX WPI; 2005-178323/19.  
DR N-PSDB; ADY72730.  
XX  
XX Method for preparing mutant gene by using, as mutagen, radiation selected  
PT from UVA (ultra violet a), UVB (ultra violet b), UVC (ultra violet c), x-  
PT ray, VUV (vacuum ultra violet) and ultrasoft x-ray.  
XX  
PS Disclosure; Fig 1; 20pp; Korean.  
XX  
CC The invention relates to a method for preparing a mutant gene by  
CC radiation is provided, to mutate a specific gene selectively by direct  
CC radiation. Therefore, it can be useful for improving genes of industrial  
CC enzymes and developing enzymes having novel properties. The present  
CC sequence represents a dextranucrase mutant DNA.  
XX  
SQ Sequence 1477 AA;  
  
Query Match 60.0%; Score 51; DB 9; Length 1477;  
Best Local Similarity 64.3%; Pred. No. 52;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 QWNGESEKPYDDHL 14  
Db 397 QWNMSSEDPKNDHL 410  
||| ||| :|||  
  
RESULT 43  
AAM86301  
ID AAM86301 standard; protein; 102 AA.  
XX  
AC AAM86301;  
XX  
DT 07-NOV-2001 (first entry)  
XX  
DE Human immune/haematopoietic antigen SEQ ID NO:13894.  
XX  
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
KW cytostatic; gene therapy; vaccine; metastasis.  
XX  
OS Homo sapiens.  
XX  
PN WO200157182-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US001354.  
XX  
PR 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.  
PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 11-JUL-2000; 2000US-0216880P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 14-JUL-2000; 2000US-0217496P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225447P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225758P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 18-AUG-2000; 2000US-0226279P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226686P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230437P.

XX Kang HG, Kim CY, Kim DM, Seo ES;  
XX WPI; 2005-178323/19.  
DR N-PSDB; ADY72730.  
XX  
XX Method for preparing mutant gene by using, as mutagen, radiation selected  
PT from UVA (ultra violet a), UVB (ultra violet b), UVC (ultra violet c), x-  
PT ray, VUV (vacuum ultra violet) and ultrasoft x-ray.  
XX  
PS Disclosure; Fig 1; 20pp; Korean.  
XX  
CC The invention relates to a method for preparing a mutant gene by  
CC radiation is provided, to mutate a specific gene selectively by direct  
CC radiation. Therefore, it can be useful for improving genes of industrial  
CC enzymes and developing enzymes having novel properties. The present  
CC sequence represents a dextranucrase mutant DNA.  
XX  
SQ Sequence 1477 AA;  
  
Query Match 60.0%; Score 51; DB 9; Length 1477;  
Best Local Similarity 64.3%; Pred. No. 52;  
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
  
QY 1 QWNGESEKPYDDHL 14  
Db 397 QWNMSSEDPKNDHL 410  
||| ||| :|||  
  
RESULT 42  
ADY72696  
ID ADY72696 standard; protein; 1477 AA.  
XX  
AC ADY72696;  
XX  
DT 19-MAY-2005 (first entry)  
XX  
DE Mutant dextranucrase.  
XX  
KW mutein; dextranucrase; food; cosmetic; pharmaceutical; enzyme.  
XX  
OS Unidentified.  
XX  
PN KR2004092250-A.  
XX  
PD 03-NOV-2004.  
XX  
PF 25-APR-2003; 2003KR-00026495.  
XX  
PR 25-APR-2003; 2003KR-00026495.  
XX  
PA (KIMD/) KIM D M.  
XX  
PI Cho GS, Kang HG, Kim DM, Seo ES;  
XX  
XX WPI; 2005-178324/19.  
DR N-PSDB; ADY72695.  
XX  
XX Mutant gene of dextranucrase, recombinant vector comprising the same  
PT gene, microorganism transformed with the same vector, and dextranucrase  
PT produced therefrom.  
XX  
XX Claim 5; SEQ ID NO 2; 20pp; Korean.  
XX  
XX The invention relates to a mutant gene of dextranucrase, a recombinant  
CC vector comprising the same gene, a microorganism transformed with the  
CC same vector, and dextranucrase produced therefrom are provided, thereby  
CC improving the production yield of dextran having a high ratio of branched  
CC bond which is useful in medicines, foods and cosmetics. The present  
CC sequence represents the mutant dextranucrase.  
XX  
SQ Sequence 1477 AA;

PR	06-SEP-2000;	2000US-02304380P
PR	08-SEP-2000;	2000US-02311242P
PR	08-SEP-2000;	2000US-02311243P
PR	08-SEP-2000;	2000US-02311244P
PR	08-SEP-2000;	2000US-02311413P
PR	08-SEP-2000;	2000US-02311414P
PR	08-SEP-2000;	2000US-02320800P
PR	08-SEP-2000;	2000US-02320810P
PR	12-SEP-2000;	2000US-02311968P
PR	14-SEP-2000;	2000US-0232397P
PR	14-SEP-2000;	2000US-0232398P
PR	14-SEP-2000;	2000US-0232399P
PR	14-SEP-2000;	2000US-02324000P
PR	14-SEP-2000;	2000US-0232401P
PR	14-SEP-2000;	2000US-02320633P
PR	14-SEP-2000;	2000US-0233064P
PR	14-SEP-2000;	2000US-0233065P
PR	21-SEP-2000;	2000US-0234223P
PR	21-SEP-2000;	2000US-0234274P
PR	25-SEP-2000;	2000US-0234997P
PR	25-SEP-2000;	2000US-0234998P
PR	26-SEP-2000;	2000US-0235484P
PR	27-SEP-2000;	2000US-0235634P
PR	27-SEP-2000;	2000US-0235836P
PR	29-SEP-2000;	2000US-0236372P
PR	29-SEP-2000;	2000US-0236376P
PR	29-SEP-2000;	2000US-0236377P
PR	29-SEP-2000;	2000US-0234998P
PR	26-SEP-2000;	2000US-0235484P
PR	29-SEP-2000;	2000US-0236369P
PR	29-SEP-2000;	2000US-0236370P
PR	02-OCT-2000;	2000US-0236802P
PR	02-OCT-2000;	2000US-0237037P
PR	02-OCT-2000;	2000US-0237038P
PR	02-OCT-2000;	2000US-0237039P
PR	02-OCT-2000;	2000US-0237040P
PR	13-OCT-2000;	2000US-0239353P
PR	13-OCT-2000;	2000US-0239937P
PR	20-OCT-2000;	2000US-0240960P
PR	20-OCT-2000;	2000US-0241121P
PR	20-OCT-2000;	2000US-0241785P
PR	20-OCT-2000;	2000US-0241786P
PR	20-OCT-2000;	2000US-0241787P
PR	20-OCT-2000;	2000US-0241808P
PR	20-OCT-2000;	2000US-0241809P
PR	20-OCT-2000;	2000US-0241826P
PR	01-NOV-2000;	2000US-0244617P
PR	08-NOV-2000;	2000US-0246474P
PR	08-NOV-2000;	2000US-0246475P
PR	08-NOV-2000;	2000US-0246476P
PR	08-NOV-2000;	2000US-0246477P
PR	08-NOV-2000;	2000US-0246478P
PR	08-NOV-2000;	2000US-0246523P
PR	08-NOV-2000;	2000US-0246524P
PR	08-NOV-2000;	2000US-0246525P
PR	08-NOV-2000;	2000US-0246526P
PR	08-NOV-2000;	2000US-0246611P
PR	08-NOV-2000;	2000US-0246613P
PR	17-NOV-2000;	2000US-0249207P
PR	17-NOV-2000;	2000US-0249208P
PR	17-NOV-2000;	2000US-0249209P
PR	17-NOV-2000;	2000US-0249210P
PR	17-NOV-2000;	2000US-0249211P
PR	17-NOV-2000;	2000US-0249212P
PR	17-NOV-2000;	2000US-0249213P
PR	17-NOV-2000;	2000US-0249214P
PR	17-NOV-2000;	2000US-0249215P
PR	17-NOV-2000;	2000US-0249216P
PR	17-NOV-2000;	2000US-0249217P
PR	17-NOV-2000;	2000US-0249218P

PR	17-NOV-2000;	2000US-02492244P.
PR	17-NOV-2000;	2000US-02492245P.
PR	17-NOV-2000;	2000US-02492644P.
PR	17-NOV-2000;	2000US-02492655P.
PR	17-NOV-2000;	2000US-02492977P.
PR	17-NOV-2000;	2000US-02492998P.
PR	17-NOV-2000;	2000US-02493000P.
PR	01-DEC-2000;	2000US-02501600P.
PR	01-DEC-2000;	2000US-02503911P.
PR	05-DEC-2000;	2000US-02510300P.
PR	05-DEC-2000;	2000US-02519889P.
PR	05-DEC-2000;	2000US-02567199P.
PR	06-DEC-2000;	2000US-02514799P.
PR	08-DEC-2000;	2000US-02518568P.
PR	08-DEC-2000;	2000US-02518689P.
PR	08-DEC-2000;	2000US-02519893P.
FR	08-DEC-2000;	2000US-02519900P.
PR	11-DEC-2000;	2000US-02540978P.
PR	05-JAN-2001;	2001US-0259678P.
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX	Rosen CA, Barash SC, Ruben SM;	
PI	WPI; 2001-483426//52.	
DR	N-PSDB; AAK59082.	
XX		
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
FT	useful for preventing, diagnosing and/or treating cancers and metastasis.	
XX		
PS	Claim 11; SEQ ID NO 13894; 3071pp + Sequence Listing; English.	
XX		
CC	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)	
CC	amino acid sequences given in AAM821170 to AAK91921. (I) have cytostatic	
CC	activity, and can be used in gene therapy and vaccine production. (I)	
CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
CC	treatment of diseases associated with inappropriate (I) expression. For	
CC	example, they may be used to treat disorders associated with decreased	
CC	expression by rectifying mutations or deletions in a patient's genome	
CC	that affect the activity of (I) by expressing inactive proteins or to	
CC	supplement the patients own production of (I). Additionally, (I)	
CC	polynucleotides may be used to produce the secreted (I), by inserting the	
CC	nucleic acids into a host cell and culturing the cell to express the	
CC	protein. (I) proteins and polynucleotides may be used to prevent,	
CC	diagnose and treat immune/haematopoietic-related diseases, especially	
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
CC	to AAK87694 represent human immune/haematopoietic antigen genomic	
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169	
CC	represent sequences used in the exemplification of the present invention	
XX		
SQ	Sequence 102 AA;	
Query Match	57.6%; Score 49; DB 4; Length 102;	
Best Local Similarity	58.3%; Pred. No. 5.5;	
Matches	7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;	
Qy	1 QWNGSEKPYDD 12	
	:::	
Dd	74 QWGGEAQRPKDD 85	
RESULT 44		
ABU50118		
ID	ABU50118 standard; protein; 105 AA.	
XX		
AC	ABU50118;	
XX		
DT	19-JUN-2003 (first entry)	
XX		
DE	Protein encoded by Prokaryotic essential gene #35645.	
XX		
KW	Antisense; prokaryotic essential gene; cell proliferation drug design.	

```

XX OS Yersinia pestis.
XX AC WO200277183-A2.
XX PD 03-OCT-2002.
XX PF 21-MAR-2002; 2002WO-US009107.
XX PR 21-MAR-2001; 2001US-00815242.
XX PR 06-SEP-2001; 2001US-00948993.
XX PR 25-OCT-2001; 2001US-0342923P.
XX PR 08-FEB-2002; 2002US-00072851.
XX PR 06-MAR-2002; 2002US-0362699P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW,
XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX PF WPI: 2003-029926/02.
XX DR N-PSDB; ACAS3988.
XX PR New antisense nucleic acids, useful for identifying proteins or screening
XX PT for homologous nucleic acids required for cellular proliferation to
XX PT isolate candidate molecules for rational drug discovery programs.
XX PS Claim 25; SEQ ID NO 78042; 1766pp; English.
XX CC The invention relates to an isolated nucleic acid comprising any one of
XX CC the 6213 antisense sequences given in the specification where expression
XX CC of the nucleic acid inhibits proliferation of a cell. Also included are:
XX CC (1) a vector comprising a promoter operably linked to the nucleic acid
XX CC encoding a polypeptide whose expression is inhibited by the antisense
XX CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
XX CC polypeptide or its fragment whose expression is inhibited by the
XX CC antisense nucleic acid; (4) an antibody capable of specifically binding
XX CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
XX CC proliferation or the activity of a gene in an operon required for
XX CC proliferation; (7) identifying a compound that influences the activity of
XX CC the gene product or that has an activity against a biological pathway
XX CC required for proliferation, or that inhibits cellular proliferation; (8)
XX CC identifying a gene required for cellular proliferation or the biological
XX CC pathway in which a proliferation-required gene or its gene product lies
XX CC or a gene on which the test compound that inhibits proliferation of an
XX CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
XX CC compound's activity; (11) a culture comprising strains in which the gene
XX CC product is overexpressed or underexpressed; (12) determining the extent
XX CC to which each of the strains is present in a culture or collection of
XX CC strains; or (13) identifying the target of a compound that inhibits the
XX CC proliferation of an organism. The antisense nucleic acids are useful for
XX CC identifying proteins or screening for homologous nucleic acids required
XX CC for cellular proliferation to isolate candidate molecules for rational
XX CC drug discovery programs, or for screening homologous nucleic acids
XX CC required for proliferation in cells other than S. aureus, S. typhimurium,
XX CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
XX CC the target prokaryotic essential genes. Note: The sequence data for this
XX CC patent did not form part of the printed specification, but was obtained
XX CC in electronic format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences
XX SQ Sequence 105 AA;

Query Match 56.5%; Score 48; DB 6; Length 105;
Best Local Similarity 53.8%; Pred. No. 8.3;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QWNGSEKPYDDH 13
Db :||| :|
3 EWNGEYVSPAEH 15

RESULT 45

```

```

ADF04672
XX ID ADF04672 standard; protein; 357 AA.
XX AC ADF04672;
XX DT 12-FEB-2004 (first entry)
XX DE Bacterial polypeptide #785.
XX KW Proteus mirabilis infection; bacterial infection; antibacterial;
XX KW immunostimulant.
XX OS Proteus mirabilis.
XX PN US6605709-B1.
XX PD 12-AUG-2003.
XX PF 05-APR-2000; 2000US-00543681.
XX PR 09-APR-1999; 99US-0128706P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Breton GL;
XX DR WPI: 2003-895291/82.
XX DR N-PSDB; ADF00500.
XX PR New Proteus mirabilis polypeptides and polynucleotides, useful as
XX PT reagents for diagnosis of bacterial disease, as components of
XX PT antibacterial vaccines, as targets for antibacterial drugs, or as
XX PT biocontrol agents for plants.
XX PS Disclosure; SEQ ID NO 4957; 870pp; English.
XX CC The invention relates to new Proteus mirabilis polypeptides and
XX CC polynucleotides. The invention also relates to antibodies against the
XX CC polypeptides, methods for producing the polypeptides, a method of
XX CC generating vaccines for immunising an individual against P. mirabilis, a
XX CC method for evaluating a compound for the ability to bind a P. mirabilis
XX CC polypeptide and a method for screening test compounds for anti-bacterial
XX CC activity. The polypeptides and polynucleotides are useful as molecular
XX CC targets for diagnosing, preventing and treating pathological conditions
XX CC resulting from bacterial infection, as reagents for diagnosis of
XX CC bacterial diseases, as components of antibacterial vaccines, as targets
XX CC for antibacterial drugs or as bio-control agents for plants. This
XX CC sequence represents a Proteus mirabilis polypeptide of the invention.
XX SQ Sequence 357 AA;

Query Match 55.9%; Score 47.5; DB 7; Length 357;
Best Local Similarity 47.4%; Pred. No. 39;
Matches 9; Conservative 1; Mismatches 2; Indels 7; Gaps 1;

QY 2 WNGSEKPK-----YDDH 13
Db :||| :|
40 WNGKDEKPKQIATLPSPYDPH 58

RESULT 46
ABU40591
XX ID ABU40591 standard; protein; 105 AA.
XX AC ABU40591;
XX DT 19-JUN-2003 (first entry)
XX DE Protein encoded by Prokaryotic essential gene #26118.
XX KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX OS Proteus sp.

```



XX WO200277183-A2.  
XX 03-OCT-2002.  
XX 21-MAR-2002; 2002WO-US009107.  
XX 21-MAR-2001; 2001US-00815242.  
XX 06-SEP-2001; 2001US-00948993.  
XX 25-OCT-2001; 2001US-0342923P.  
XX 08-FEB-2002; 2002US-00072851.  
XX 06-MAR-2002; 2002US-0362699P.  
XX (ELIT-) ELITRA PHARM INC.  
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX WPI; 2003-029926/02.  
XX N-PSDB; ACA4461.  
XX New antisense nucleic acids, useful for identifying proteins or screening  
XX for homologous nucleic acids required for cellular proliferation to  
XX isolate candidate molecules for rational drug discovery programs.  
XX Claim 25; SEQ ID NO 68515; 1766pp; English.  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation or the activity of a gene in an operon required for  
XX the gene product; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or the biological  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational  
XX drug discovery programs, or for screening homologous nucleic acids  
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
XX the target prokaryotic essential genes. Note: The sequence data for this  
XX patent did not form part of the printed specification, but was obtained  
XX in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 105 AA;  
Query Match 55.3%; Score 47; DB 6; Length 105;  
Best Local Similarity 53.8%; Pred. No. 12;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 QWNGSEKPYDDH 13  
:||||| :|  
Db 3 EWNGEYISPYAEH 15  
RESULT 47  
ADF08022  
ID ADF08022 standard; protein; 106 AA.

XX ADF08022;  
XX 12-FEB-2004 (first entry)  
XX Bacterial polypeptide #4135.  
XX Proteus mirabilis infection; bacterial infection; antibacterial;  
XX immunostimulant.  
XX Proteus mirabilis.  
XX US6605709-B1.  
XX 12-AUG-2003.  
XX 05-APR-2000; 2000US-00543681.  
XX 09-APR-1999; 99US-0128706P.  
XX (GENO-) GENOME THERAPEUTICS CORP.  
XX Breton GL;  
XX WPI; 2003-895291/82.  
XX N-PSDB; ADF03850.  
XX New Proteus mirabilis polypeptides and polynucleotides, useful as  
XX reagents for diagnosis of bacterial disease, as components of  
XX antibacterial vaccines, as targets for antibacterial drugs, or as  
XX biocontrol agents for plants.  
XX Disclosure; SEQ ID NO 8307; 870pp; English.  
XX The invention relates to new Proteus mirabilis polypeptides and  
XX polynucleotides. The invention also relates to antibodies against the  
XX polypeptides, methods for producing the polypeptides, a method of  
XX generating vaccines for immunising an individual against *P. mirabilis*, a  
XX method for evaluating a compound for the ability to bind a *P. mirabilis*, a  
XX polypeptide and a method for screening test compounds for anti-bacterial  
XX activity. The polypeptides and polynucleotides are useful as molecular  
XX targets for diagnosing, preventing and treating pathological conditions  
XX resulting from bacterial infection, as reagents for diagnosis of  
XX bacterial diseases, as components of antibacterial vaccines, as targets  
XX for antibacterial drugs or as bio-control agents for plants. This  
XX sequence represents a Proteus mirabilis polypeptide of the invention.  
XX  
XX Sequence 106 AA;  
Query Match 55.3%; Score 47; DB 7; Length 106;  
Best Local Similarity 53.8%; Pred. No. 12;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 QWNGSEKPYDDH 13  
:||||| :|  
Db 4 EWNGEYISPYAEH 16  
RESULT 48  
ABR63236  
ID ABR63236 standard; protein; 1149 AA.  
XX ABR63236;  
XX 27-AUG-2003 (first entry)  
XX Glucansucrase sequence from strain Lb33.  
XX Glucan; glucosyltransferase activity; thickener; prebiotic;  
XX bioactive agent; anti-corrosion agent; foodstuff; gastrointestinal tract.  
XX Lactobacillus fermentum.  
XX

```
PN WO2003008618-A2.
XX
XX 30-JAN-2003.
XX
XX 22-JUL-2002; 2002WO-NL000495.
XX
XX 20-JUL-2001; 2001EP-00202752.
XX
XX 25-JUL-2001; 2001EP-00202841.
XX
XX (NEDE ) NEDERLANDSE ORG TOEGEPAST.
XX
XX Van Geel- Schutten GH;
XX
XX WPI; 2003-289780/28.
XX
XX N-PSDB; ACC84452.
XX
XX Novel glucan produced by glucosyltransferase activity of lactic acid
XX bacterium on sucrose substrate, and having backbone consisting of alpha
XX (1,3)- and alpha (1,6)-linked anhydroglucose units, useful as thickener.
XX
XX Claim 11; Fig 1; Sipp; English.
XX
XX The present invention relates to glucan capable of being produced by
XX glucosyltransferase activity of a lactic acid bacterium on a sucrose
XX substrate. The method is useful as a thickener, as a prebiotic and as a
XX bioactive agent and as an anti-corrosion agent. The glucan can be
XX incorporated in foodstuffs such as beverages, sauces, dressings, dairy
XX products. The glucan is useful as anticorrosion agent, e.g. for the
XX protection of ship hulls. It can also be incorporated in nutritional or
XX pharmaceutical compositions intended for improving the condition of the
XX gastrointestinal tract. The present sequence represents a sequence of the
XX glucansucrase gene
XX
XX Sequence 1149 AA;
SQ
Query Match 55.3%; Score 47; DB 6; Length 1149;
Best Local Similarity 64.3%; Pred. No. 1.8e+02;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 QWNGSEKPYDDHL 14
DB 11 QWNKTSDDVNDHL 24
RESULT 49
ARM80111
ID AAM80111 standard; protein; 138 AA.
XX
XX AAM80111;
XX
XX 06-NOV-2001 (first entry)
XX
XX Human protein SEQ ID NO 3757.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
XX
XX WO200157190-A2.
XX
XX 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US0004098.
XX
XX 03-FEB-2000; 2000US-00496914.
XX
XX 27-APR-2000; 2000US-00560875.
XX
XX 20-JUN-2000; 2000US-00598075.
XX
XX 19-JUL-2000; 2000US-00620325.
XX
XX 01-SEP-2000; 2000US-00654936.
XX
XX 15-SEP-2000; 2000US-00663561.
XX
PR 20-OCT-2000; 2000US-00693325.
XX
XX 30-NOV-2000; 2000US-00728422.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
XX Ma Y, Zhao QA, Wang D, Zhang J, Zhang J, Ren F, Chen R, Wang ZW;
XX Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
XX
XX N-PSDB; AAK53244.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
XX in diagnosis and gene therapy.
XX
XX Claim 20; Page 427; 5221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
XX (AAK52582) and 3686 (AAM80020) are omitted as the relevant pages from the
XX sequence listing were missing at the time of publication
XX
XX Sequence 138 AA;
SQ
Query Match 54.1%; Score 46; DB 4; Length 138;
Best Local Similarity 61.5%; Pred. No. 24;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 WNGSEKPYDDHL 14
DB 65 WKGLSEKVDQHL 77
RESULT 50
ABP38135
ID ABP38135 standard; protein; 504 AA.
XX
XX ABP38135;
XX
XX 24-JUL-2002 (first entry)
XX
XX Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:2980.
XX
XX Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX antibacterial; gene therapy.
XX
XX Staphylococcus epidermidis.
XX
XX US6380370-B1.
XX
XX 30-APR-2002.
XX
XX 13-AUG-1998; 98US-00134001.
XX
XX 14-AUG-1997; 97US-0055779P.
XX
XX 08-NOV-1997; 97US-0064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI; 2002-381255/41.
XX
XX N-PSDB; ABN90680.
XX
```

1990

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:30:52 ; Search time 42 Seconds  
(without alignments)  
32.072 Million cell updates/sec

Title: US-10-797-821-28

Perfect score: 85

Sequence: 1 QWNGSEKPYDDHL 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : PIR 80.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85	100.0	1592	2 A38175	glucosyltransferase
2	67	78.8	1375	2 J03345	dextranase (EC
3	67	78.8	1475	2 B31135	grfB protein precu
4	51	60.0	1508	2 T31098	probable dextranu
5	48	56.5	105	2 AH0014	transcription repr
6	45	52.9	105	2 E82047	met repressor VC26
7	44	51.8	170	2 T35937	hypothetical prote
8	44	51.8	416	2 T08267	hypothetical prote
9	43	50.6	202	2 JC4635	tumor-associated L
10	43	50.6	202	2 JC6205	cell surface antig
11	43	50.6	347	2 T06584	probable DNA-bindi
12	43	50.6	1290	2 JC5474	dextranase (EC
13	43	50.6	4767	2 T31345	hypothetical prote
14	42	49.4	105	1 RGECHJ	Met regulon regula
15	42	49.4	105	2 AB0938	repressor of the m
16	42	49.4	105	2 C86084	repressor of all m
17	42	49.4	105	2 C91237	repressor protein
18	42	49.4	105	2 A23081	regulatory protein
19	42	49.4	125	2 H83068	hypothetical prote
20	42	49.4	424	2 B95201	3-hydroxy-3-methyl
21	42	49.4	424	2 H98067	hydroxymethylgluta
22	42	49.4	573	2 S29623	hydroxymethylgluta
23	42	49.4	602	2 T35782	probable secreted
24	42	49.4	1429	2 S06434	homeotic protein 1
25	41.5	48.8	447	2 T13204	hypothetical prote
26	41	48.2	225	2 T35679	hypothetical prote
27	41	48.2	389	2 A87602	conserved hypothet
28	41	48.2	813	2 E85135	hypothetical prote
29	40	47.1	179	2 T32925	hypothetical prote

30	40	47.1	346	2 A03308	dnaK-type molecula
31	40	47.1	457	2 AF3321	alpha-ketoglutarat
32	40	47.1	532	2 JC6170	GATA-transcription
33	40	47.1	863	2 JC7537	beta-N-acetylgluco
34	40	47.1	1001	2 S66704	hypothetical prote
35	39.5	46.5	1123	2 A44766	defective chorion-
36	39	45.9	304	2 E95909	hypothetical prote
37	39	45.9	339	1 R6B113	ribosomal protein
38	39	45.9	400	2 A45545	major merozoite su
39	39	45.9	449	2 T31546	hypothetical prote
40	39	45.9	464	2 T28662	hypothetical prote
41	39	45.9	492	2 A97200	membrane associate
42	39	45.9	517	2 I40798	cellulase (EC 3.2.
43	39	45.9	589	2 T12820	hypothetical prote
44	39	45.9	625	2 G72420	oligopeptide ABC t
45	39	45.9	657	2 AH2913	acetyl-coenzyme A
46	39	45.9	657	2 C97688	acs(acetyl-CoA syn
47	39	45.9	1060	2 S63993	acrosomal protein
48	39	45.9	1145	2 T33606	hypothetical prote
49	39	45.9	1631	1 SAZOK1	major merozoite su
50	39	45.9	1639	2 S05603	probable major sur
51	39	45.9	1640	2 A24594	hypothetical prote
52	38.5	45.3	535	2 S76564	hypothetical prote
53	38.5	45.3	614	2 T30973	hypothetical prote
54	38.5	45.3	616	2 C86239	protein T10024.21
55	38	44.7	80	2 C91287	hypothetical prote
56	38	44.7	156	2 T41316	peroxisomal membra
57	38	44.7	198	2 B97088	amidase, related t
58	38	44.7	224	2 C46357	env polyprotein -
59	38	44.7	247	2 B45711	transmembrane prot
60	38	44.7	288	2 T47811	hypothetical prote
61	38	44.7	338	2 A86412	protein Flk33.10 [
62	38	44.7	350	2 T10498	UDPglucose 4-epime
63	38	44.7	365	2 AG2970	periplasmic mannit
64	38	44.7	365	2 C98312	periplasmic mannit
65	38	44.7	383	2 T00674	hypothetical prote
66	38	44.7	387	2 T33413	hypothetical prote
67	38	44.7	404	2 T25420	hypothetical prote
68	38	44.7	461	2 S57713	hypothetical prote
69	38	44.7	465	2 A38079	probable mannosyl
70	38	44.7	495	2 H72297	gamma-aminobutyric
71	38	44.7	506	2 C64377	TRK system potassi
72	38	44.7	600	2 T32885	hypothetical prote
73	38	44.7	859	1 VCLJ22	env polyprotein pr
74	38	44.7	859	1 VCLJ21	env polyprotein pr
75	38	44.7	859	1 VCLJ22	env polyprotein pr
76	38	44.7	859	1 VCLJ23	env polyprotein pr
77	38	44.7	859	1 VCLJ24	env polyprotein pr
78	38	44.7	859	1 VCLJ25	env polyprotein pr
79	38	44.7	859	1 VCLJ26	env polyprotein pr
80	38	44.7	860	1 VCLJ27	env polyprotein pr
81	38	44.7	888	2 T51593	GTP-binding regula
82	38	44.7	901	2 T01135	probable GTP-bindi
83	38	44.7	1010	2 D72203	hypothetical prote
84	38	44.7	1301	2 S18118	alpha-amylase - Al
85	38	44.7	1305	2 A40879	phospholipase C (E
86	38	44.7	1312	1 B40879	phospholipase C (E
87	37.5	44.1	416	2 S65110	chitinase (EC 3.2.
88	37.5	44.1	3083	2 AH2493	hypothetical prote
89	37	43.5	73	2 F90221	LSU ribosomal prot
90	37	43.5	105	2 T24359	hypothetical prote
91	37	43.5	128	2 C48552	hypothetical prote
92	37	43.5	158	2 T49419	hypothetical prote
93	37	43.5	165	2 A90169	hypothetical prote
94	37	43.5	193	2 H69938	conserved hypothet
95	37	43.5	209	2 D86107	hypothetical prote
96	37	43.5	209	2 G91266	hypothetical prote
97	37	43.5	209	2 S56350	hypothetical 23.4K
98	37	43.5	234	2 T46941	conjugal transfer
99	37	43.5	234	2 S37463	regulatory protein
100	37	43.5	271	2 H97230	alpha/beta superfa
101	37	43.5	351	2 T08446	hypothetical prote
102	37	43.5	367	2 C44102	di-N-acetylchitobi

103	37	43.5	377	2	AE1991	hypothetical prote	176	36	42.4	1	T01236	serine/threonine-s
104	37	43.5	437	2	A97111	Fe-S oxidoreductas	177	36	42.4	2	AE2713	cytosol aminopepti
105	37	43.5	445	2	T33617	hypothetical prote	178	36	42.4	2	B97495	hypothetical prote
106	37	43.5	455	2	T00856	pectate lyase [EC	179	36	42.4	2	T06029	hypothetical prote
107	37	43.5	469	2	H83591	S-adenosyl-L-homoc	180	36	42.4	2	S58224	oestrogen receptor
108	37	43.5	471	2	T10189	hypothetical prote	181	36	42.4	2	S44188	alpha-glucosidase
109	37	43.5	500	2	H70629	probable AMINOPEPT	182	36	42.4	2	G97133	phage terminase-li
110	37	43.5	531	2	A84471	En/Spm-like transp	183	36	42.4	2	S52629	catechol oxidase (
111	37	43.5	550	2	D64354	guanine tRNA-ribos	184	36	42.4	2	B70976	glutamine-fructose
112	37	43.5	554	2	H83053	glucose-6-phosphat	185	36	42.4	2	B95878	probable adenylate
113	37	43.5	599	1	WNBPT9	gene 9 protein - p	186	36	42.4	2	T52064	dnau-like protein
114	37	43.5	611	2	F75095	probable asparagin	187	36	42.4	1	JX0336	succinate dehydrog
115	37	43.5	650	2	H84356	long-chain fatty-a	188	36	42.4	2	T04876	hypothetical prote
116	37	43.5	748	2	L48744	semaphorin A - mou	189	36	42.4	2	S75636	sensory transducti
117	37	43.5	749	2	G01856	semaphorin V - hum	190	36	42.4	2	A10062	conserved hypochet
118	37	43.5	750	2	T04980	hypothetical prote	191	36	42.4	2	A10062	interleukin-4 rece
119	37	43.5	750	2	F83003	hypothetical prote	192	36	42.4	2	JC4520	hypothetical prote
120	37	43.5	751	2	D72338	(p)ppGpp synthetas	193	36	42.4	2	T12535	probable C2H2-type
121	37	43.5	756	2	D85079	hypothetical prote	194	36	42.4	2	T02417	alpha-amyase [EC
122	37	43.5	757	2	T02561	probable cellulose	195	36	42.4	2	A42466	calmodulin-binding
123	37	43.5	819	2	S15169	ferric-pseudobacti	196	36	42.4	2	T51257	calmodulin-binding
124	37	43.5	844	1	A26528	penicillin amidase	197	36	42.4	2	T50928	protein F1E22.12 l
125	37	43.5	894	2	B87990	protein Y6B3A.1 [l	198	36	42.4	2	A96682	1-phosphatidylinos
126	37	43.5	894	2	T27321	inter-alpha-trypsi	199	36	42.4	2	A53970	glucosyltransferas
127	37	43.5	911	2	A39967	reverse transcript	200	36	42.4	2	S22737	zinc finger protei
128	37	43.5	964	2	T01860	neural cell adhesi	201	36	42.4	2	T14273	ubiquitin-protein
129	37	43.5	1260	1	S05479	protein kinase yak	202	36	42.4	2	T30554	hypothetical prote
130	37	43.5	1457	2	T14577	hypothetical prote	203	36	42.4	2	T26063	aggrecon precursor
131	37	43.5	4131	2	T21085	hypothetical prote	204	36	42.4	1	A39086	367K tegument prot
132	36.5	42.9	112	2	T13322	hypothetical prote	205	36	42.4	2	WZBBE6	tegument protein 2
133	36.5	42.9	143	2	AH0071	hypothetical prote	206	36	42.4	2	T42567	C2 protein - tobac
134	36.5	42.9	188	2	G86690	hypothetical prote	207	35.5	41.8	2	C42452	conserved hypochet
135	36.5	42.9	270	2	D69456	alkaline serine pr	208	35.5	41.8	2	B75432	hypothetical prote
136	36.5	42.9	321	2	T32320	hypothetical prote	209	35.5	41.8	2	T37202	hypothetical prote
137	36.5	42.9	327	2	T32321	conserved hypochet	210	35.5	41.8	2	T01926	hypothetical prote
138	36.5	42.9	393	2	C64613	E1 protein - human	211	35.5	41.8	2	T08187	hypothetical prote
139	36.5	42.9	605	2	S36469	hypothetical prote	212	35.5	41.8	2	G70556	probable trpE prot
140	36.5	42.9	1032	2	F65071	probable oxidoredu	213	35.5	41.8	2	T45254	probable anthranil
141	36.5	42.9	1032	2	C85943	probable oxidoredu	214	35.5	41.8	2	AD2324	amidase enhancer [
142	36.5	42.9	1032	2	G91097	probable oxidoredu	215	35.5	41.8	2	F69837	asparagine synthas
143	36.5	42.9	3351	2	T13812	lipophorin - fruit	216	35.5	41.8	2	B81280	probable periplasm
144	36	42.4	56	2	E87419	hypothetical prote	217	35.5	41.8	2	B83404	hypothetical prote
145	36	42.4	105	2	A64060	regulatory protein	218	35.5	41.8	2	T21970	hypothetical prote
146	36	42.4	152	2	F71973	hypothetical prote	219	35.5	41.8	2	A49674	flightless-I homol
147	36	42.4	174	2	A81417	probable integral	220	35.5	41.8	2	T00020	bacterial blight-r
148	36	42.4	183	2	S29088	ubiquitin-protein	221	35.5	41.8	2	T43431	alpha-glucan synth
149	36	42.4	186	2	D64533	hypothetical prote	222	35.5	41.8	2	T00327	polyprotein - infe
150	36	42.4	188	2	T33886	hypothetical prote	223	35	41.2	70	B82895	hypothetical prote
151	36	42.4	223	2	D75515	hypothetical prote	224	35	41.2	2	AB1683	conserved hypochet
152	36	42.4	248	2	C90940	hypothetical prote	225	35	41.2	2	F75250	probable deoxycyti
153	36	42.4	248	2	G85788	hypothetical prote	226	35	41.2	1	ZZSYN2	nodulin-24 - soybe
154	36	42.4	248	2	F64938	hypothetical prote	227	35	41.2	2	D82795	phage-related endo
155	36	42.4	286	2	S16423	general stress pro	228	35	41.2	2	G95343	hypothetical prote
156	36	42.4	317	2	D95044	diphosphomevalonat	229	35	41.2	2	AE2808	succinoglycan bios
157	36	42.4	336	2	S70708	site-specific DNA-	230	35	41.2	2	B86576	CT482 hypochetical
158	36	42.4	344	2	C97914	diphosphomevalonat	231	35	41.2	2	G91207	probable replicase
159	36	42.4	347	2	A86549	polymorphic outer	232	35	41.2	2	A65172	hypothetical prote
160	36	42.4	356	2	F82629	phage-related prot	233	35	41.2	2	C97587	succinoglycan bios
161	36	42.4	358	1	PASPC	fructose-bisphosph	234	35	41.2	2	T03046	hypothetical prote
162	36	42.4	359	2	C82626	phage-related prot	235	35	41.2	2	AB3872	hypothetical prote
163	36	42.4	369	2	F83828	hypothetical prote	236	35	41.2	2	T34951	hypothetical prote
164	36	42.4	381	2	S69795	UTP-hexose-1-phosp	237	35	41.2	2	A69952	conserved hypochet
165	36	42.4	388	2	PN0476	1-aminocyclopropan	238	35	41.2	2	T27773	hypothetical prote
166	36	42.4	391	2	PN0477	1-aminocyclopropan	239	35	41.2	2	A59369	ubiquinol-cytochro
167	36	42.4	396	2	S56496	prophage P4 integr	240	35	41.2	2	S10641	endA protein - Str
168	36	42.4	399	2	T19115	hypothetical prote	241	35	41.2	2	F95229	DNA-entry nuclease
169	36	42.4	410	2	C84205	hypothetical prote	242	35	41.2	2	A99094	deoxyribonuclease
170	36	42.4	415	2	T09085	fructose-bisphosph	243	35	41.2	2	G69837	phoeyone synthase
171	36	42.4	421	1	S26605	myb-related protei	244	35	41.2	2	A84429	probable S-locus g
172	36	42.4	428	1	B32804	GTP-binding protei	245	35	41.2	2	IS3161	alpha 2-adrenergic
173	36	42.4	430	1	A27655	adenosylhomocyste	246	35	41.2	2	AC1599	tRNA isopentenylp
174	36	42.4	434	2	AD0197	NADH2 dehydrogenas	247	35	41.2	2	E86861	methionyl-tRNA for
175	36	42.4	447	2	T06004	1-aminocyclopropan	248	35	41.2	2	A86054	probable replicase

249	35	41.2	337	2	A24330	regulatory protein	322	35	41.2	928	1	VGEMC
250	35	41.2	354	2	D31751	protein kinase cat	323	35	41.2	943	2	F70873
251	35	41.2	362	2	AG3428	GPMannose 4,6-deh	324	35	41.2	947	2	S08243
252	35	41.2	362	2	A24710	p1fC protein - Esc	325	35	41.2	980	2	T27342
253	35	41.2	366	2	A84209	hypothetical prote	326	35	41.2	982	1	VCLJVS
254	35	41.2	371	2	S46436	porin foma precurs	327	35	41.2	983	1	E45390
255	35	41.2	371	2	AD2802	hypothetical prote	328	35	41.2	984	2	E70406
256	35	41.2	376	2	D29826	hypothetical prote	329	35	41.2	1002	2	T02488
257	35	41.2	376	2	S31751	protein kinase cat	330	35	41.2	1026	2	A49750
258	35	41.2	379	2	E97581	periplasmic bindin	331	35	41.2	1122	2	B26427
259	35	41.2	385	2	G70477	hypothetical prote	332	35	41.2	1127	2	A25018
260	35	41.2	401	2	F90288	hypothetical prote	333	35	41.2	1129	2	D84022
261	35	41.2	408	2	G83406	conserved hypotcet	334	35	41.2	1165	2	T30731
262	35	41.2	414	2	F29826	hypothetical prote	335	35	41.2	1176	2	C26427
263	35	41.2	416	2	D90042	fmaA protein limpo	336	35	41.2	1187	2	C84568
264	35	41.2	427	2	T09314	sulfite reductase	337	35	41.2	1208	2	S17286
265	35	41.2	430	2	S77393	glycoprotein - hum	338	35	41.2	1218	2	A26588
266	35	41.2	430	2	S20187	cell division prot	339	35	41.2	1218	2	A26427
267	35	41.2	433	2	T40763	WP1 protein precu	340	35	41.2	1222	2	G59100
268	35	41.2	435	2	F84674	adenosylhomocystei	341	35	41.2	1310	2	AD1380
269	35	41.2	436	2	D69270	probable AAA-type	342	35	41.2	1398	2	T20434
270	35	41.2	437	2	S59831	ferredoxin-nitrite	343	35	41.2	1518	2	A44811
271	35	41.2	449	2	S50546	END13 protein - ye	344	35	41.2	1571	2	T14155
272	35	41.2	450	2	S72223	adenosylhomocystei	345	35	41.2	1679	2	T30271
273	35	41.2	473	2	T06167	alpha-2B-adrenergi	346	35	41.2	1871	2	D96698
274	35	41.2	479	1	JN0715	beta-fructofuranos	347	35	41.2	1894	2	T02155
275	35	41.2	479	1	JN0890	3-phytase (EC 3.1.	348	35	41.2	1927	2	S01168
276	35	41.2	481	2	S04605	acid phosphatase {	349	35	41.2	2319	2	A47004
277	35	41.2	484	2	S11003	glycinin G3 - soyb	350	35	41.2	2344	2	T41590
278	35	41.2	485	1	FWSVG1	glycinin G1 precu	351	35	41.2	2405	2	T08164
279	35	41.2	485	2	S11002	glycinin chain A2B	352	35	41.2	2825	2	T14271
280	35	41.2	495	1	FWSYG2	glycinin G2 precu	353	34.5	40.6	85	2	F71333
281	35	41.2	495	2	S10851	glycinin chain Ala	354	34.5	40.6	188	2	D89819
282	35	41.2	503	2	S55589	D-nopaline dehydro	355	34.5	40.6	264	2	S67861
283	35	41.2	514	2	D89775	hypothetical prote	356	34.5	40.6	360	2	T71444
284	35	41.2	525	2	AF2950	GGBEF family prote	357	34.5	40.6	378	2	C81824
285	35	41.2	525	2	E98332	hypothetical prote	358	34.5	40.6	416	2	A81053
286	35	41.2	554	2	A91250	probable portal pr	359	34.5	40.6	445	2	C69233
287	35	41.2	557	2	S21596	extracellular prot	360	34.5	40.6	475	2	S54989
288	35	41.2	558	2	I48073	glucose-6-phosphat	361	34.5	40.6	475	2	S54993
289	35	41.2	568	2	I39411	AF-9 protein - hum	362	34.5	40.6	475	2	S54994
290	35	41.2	571	1	UMFF	period clock prote	363	34.5	40.6	475	2	S54996
291	35	41.2	580	1	S33743	aspartate-tRNA lig	364	34.5	40.6	475	2	S54998
292	35	41.2	584	2	T14631	glucose-6-phosphat	365	34.5	40.6	505	1	S05380
293	35	41.2	589	2	AC1446	probable portal pr	366	34.5	40.6	530	2	A84522
294	35	41.2	609	2	S28283	hypothetical prote	367	34.5	40.6	554	2	T49643
295	35	41.2	633	2	D72284	oligopeptide ABC t	368	34.5	40.6	614	2	A88466
296	35	41.2	642	2	S55521	beta-fructofuranos	369	34.5	40.6	747	2	AD2015
297	35	41.2	642	2	H81868	probable virulence	370	34.5	40.6	1054	2	B38919
298	35	41.2	644	2	T33730	hypothetical prote	371	34.5	40.6	1057	2	S09112
299	35	41.2	651	2	T12083	beta-fructofuranos	372	34.5	40.6	1657	2	A54854
300	35	41.2	663	2	F90291	hypothetical prote	373	34	40.0	45	2	C84154
301	35	41.2	666	2	I58169	semaphorin III - m	374	34	40.0	92	2	AI3345
302	35	41.2	689	1	A40088	beta-adrenergic-re	375	34	40.0	99	2	E97014
303	35	41.2	689	1	A53791	beta-adrenergic-re	376	34	40.0	100	2	S15148
304	35	41.2	689	1	I56531	conserved hypotcet	377	34	40.0	125	2	H70399
305	35	41.2	689	2	AD0336	transketolase (EC	378	34	40.0	130	2	T47540
306	35	41.2	700	2	JC71146	hypothetical prote	379	34	40.0	130	2	A75604
307	35	41.2	725	2	T08989	hypothetical prote	380	34	40.0	139	2	T18241
308	35	41.2	732	2	AI2158	hypothetical prote	381	34	40.0	151	2	D83552
309	35	41.2	771	2	D49423	semaphorin III pre	382	34	40.0	152	2	S57423
310	35	41.2	771	2	G71409	probable replicati	383	34	40.0	154	2	S76882
311	35	41.2	772	2	A49069	collapsin - chicke	384	34	40.0	172	2	AD1906
312	35	41.2	772	2	I48747	semaphorin D - mou	385	34	40.0	174	2	S08535
313	35	41.2	776	2	T02584	probable protein k	386	34	40.0	174	2	A70828
314	35	41.2	779	2	D85086	hypothetical prote	387	34	40.0	174	2	S47069
315	35	41.2	781	2	T00945	hypothetical prote	388	34	40.0	183	2	S65815
316	35	41.2	814	1	CZCLEM	cellulase (EC 3.2.	389	34	40.0	184	2	A37201
317	35	41.2	828	2	E71417	hypothetical prote	390	34	40.0	189	2	B29795
318	35	41.2	853	2	S74457	ferrichrome-iron r	391	34	40.0	189	2	H83870
319	35	41.2	853	2	T29736	hypothetical prote	392	34	40.0	189	2	D82151
320	35	41.2	890	2	T34243	hypothetical prote	393	34	40.0	190	2	D80206
321	35	41.2	908	2	T22376	hypothetical prote	394	34	40.0	193	2	D90206

glycoprotein B pre  
aconitase hydratase  
beta-glucosidase {  
hypothetical prote  
env polypeptide pr  
DMSO reductase cha  
hypothetical prote  
beta-galactosidase  
period clock prote  
circadian rhythm p  
ATP-dependent nucl  
probable DNA-direc  
period clock prote  
hypothetical prote  
period clock prote  
period clock prote  
period clock prote  
hypothetical prote  
glycosidase homolo  
hypothetical prote  
glucosyltransferase  
zinc finger protei  
surface protein -  
probable DNA polym  
DNA-directed DNA p  
beta-glucosidase c  
coagulation factor  
probable sensor-li  
dynein alpha heavy  
poc4 protein, stre  
hypothetical prote  
conserved hypotcet  
GumL protein - Xan  
hypothetical prote  
3-oxoacyl-(acyl)-ca  
3-oxoacyl-(acyl)-ca  
lysyl endopeptidas  
reverse transcript  
reverse transcript  
reverse transcript  
reverse transcript  
transcription fact  
hypothetical prote  
hypothetical prote  
protein B0244.2 [i  
hypothetical prote  
hypothetical prote  
hypothetical prote  
Ras GTPase activat  
hypothetical prote  
hypothetical prote  
uncharacterized sm  
hypothetical prote  
hypothetical prote  
actin depolymerizi  
NADH2 dehydrogenas  
lg gamma-1 chain C  
cysteine proteinas  
hypothetical prote  
ubiquitin-conjugat  
myelin basic prote  
hypothetical prote  
finger protein HZF  
RNA-directed DNA p  
aminoglycoside N6'  
circumsporozoite p  
hypothetical prote  
hypothetical prote  
imidazoleglycerol-

395	34	40.0	199	2	T05283	hypothetical prote	468	472	2	S39719	spore coat polysac
396	34	40.0	202	2	A53399	l6 antigen - mouse	469	475	2	S47861	mei-1 protein - Ca
397	34	40.0	204	2	T50321	probable 5-formylt	470	481	2	T46606	cell cycle control
398	34	40.0	206	2	T27766	hypothetical prote	471	483	2	T16926	hypothetical prote
399	34	40.0	211	2	T42301	hypothetical prote	472	503	2	B38745	cell adhesion mole
400	34	40.0	215	2	S37458	self-incompatibili	473	507	1	A43387	polymerase-associa
401	34	40.0	215	2	S66861	probable membrane	474	507	2	JQ1929	phosphoprotein - r
402	34	40.0	230	2	F71873	hypothetical prote	475	511	2	T16279	hypothetical prote
403	34	40.0	233	2	S50960	probable membrane	476	523	2	B96835	CAD ATPase (AAAL),
404	34	40.0	261	2	T40773	hypothetical prote	477	525	2	T47409	hypothetical prote
405	34	40.0	261	2	S09647	gentamycin 3'-N-ac	478	526	1	GVNSG	spike glycoprotein
406	34	40.0	265	2	S73788	hypothetical prote	479	529	2	T14947	hypothetical prote
407	34	40.0	266	2	T43453	hypothetical prote	480	531	2	E95358	probable L-sorbose
408	34	40.0	276	2	T19845	hypothetical prote	481	574	2	T41068	hypothetical prote
409	34	40.0	278	2	H95926	conserved hypothet	482	599	1	WMBP9Z	gene 9 protein - p
410	34	40.0	280	1	FRSABA	epidermolytic toxi	483	633	2	C70358	hydrogenase (EC 1.
411	34	40.0	282	2	D72679	hypothetical prote	484	648	2	T27412	hypothetical prote
412	34	40.0	283	2	T26173	hypothetical prote	485	649	2	B96729	hypothetical prote
413	34	40.0	283	1	T15229	hypothetical prote	486	665	1	A42792	succinate dehydrog
414	34	40.0	285	1	YXECIC	isochorismatase (E	487	671	2	AI0816	conserved hypothet
415	34	40.0	285	2	F85558	isochorismatase (E	488	671	2	H91045	hypothetical prote
416	34	40.0	285	2	B90708	isochorismatase (E	489	671	2	D85890	hypothetical prote
417	34	40.0	285	2	B84789	hypothetical prote	490	671	2	A65023	hypothetical prote
418	34	40.0	285	2	D84789	hypothetical prote	491	682	2	D96976	polyposphate kina
419	34	40.0	285	2	S44085	plasma membrane in	492	690	2	AG2419	hypothetical prote
420	34	40.0	286	2	AD1919	hypothetical prote	493	699	2	T09483	Cys-rich protein R
421	34	40.0	287	2	S44084	plasma membrane in	494	709	2	S51793	probable membrane
422	34	40.0	295	2	C69180	adhesion protein -	495	710	2	AC0999	cell division cont
423	34	40.0	297	2	E86701	hypothetical prote	496	732	2	A69086	cell division cont
424	34	40.0	301	2	D44355	CD44 glycoprotein	497	733	2	H69411	cdcH protein - Hal
425	34	40.0	302	2	T37326	probable transcrip	498	742	1	S47018	cell division cycl
426	34	40.0	302	2	H86271	protein F16A14.8 [	499	742	2	C84319	cell division cycl
427	34	40.0	305	2	AF1236	tRNA isopentenylp	500	745	2	T37458	VCP-like ATPase -
428	34	40.0	314	2	S35314	transcription fact	501	763	2	T49089	hypothetical prote
429	34	40.0	314	2	A44437	regenerating liver	502	765	2	G96696	protein FlN21.14 [
430	34	40.0	315	2	S27784	phenylethanolamine	503	769	2	E90158	AAA family ATPase
431	34	40.0	317	2	T32857	hypothetical prote	504	769	2	C90186	AAA family ATPase
432	34	40.0	317	2	A39935	NF-kappaB inhibito	505	769	2	S55554	male-specific leth
433	34	40.0	324	1	G1MS	Ig gamma-1 chain C	506	771	2	E84523	En/Spm-like transp
434	34	40.0	334	2	B72301	endoglyucanase - Th	507	773	2	A6720	mel-2 protein - fr
435	34	40.0	336	2	F95925	probable cell-wall	508	774	1	RRVETC	RNA-directed RNA p
436	34	40.0	337	1	A36951	UDPglucose 4-epime	509	780	1	S43859	ATPase - Sulfolobu
437	34	40.0	342	2	JC7103	brain-specific mem	510	780	2	T31548	hypothetical prote
438	34	40.0	349	2	T10330	ribonucleotide red	511	791	2	T16031	hypothetical prote
439	34	40.0	350	2	AI0139	UTP-hexose-1-phosp	512	795	2	F75154	cell division cont
440	34	40.0	357	2	T32296	hypothetical prote	513	796	2	D97065	transketolase limp
441	34	40.0	359	2	T30497	probable ribonucle	514	798	2	B71196	probable transito
442	34	40.0	365	2	JC5528	stress-activated p	515	808	2	G86208	protein F22G5.28 [
443	34	40.0	372	2	G82181	galactose-1-phosph	516	809	2	S32899	ferric-pseudobacti
444	34	40.0	379	1	A56685	UDPglucose-hexose-	517	809	2	S67665	ubiquitin-specific
445	34	40.0	381	2	F69666	iron-sulfur cofact	518	811	2	B69512	cell division cont
446	34	40.0	391	2	B97008	cyclopropane fatty	519	840	2	D75046	transitional endop
447	34	40.0	393	1	G1MSM	Ig gamma-1 chain C	520	840	2	H71114	probable cell divi
448	34	40.0	393	2	AI1312	tRNA CCA-adding en	521	847	2	T04772	hypothetical prote
449	34	40.0	393	2	B84758	probable katanin [	522	870	2	B71698	hypothetical prote
450	34	40.0	400	2	T35334	probable membrane	523	903	1	C64444	cell division cont
451	34	40.0	406	2	JX0346	alpha-1-antiprotei	524	907	2	S54353	inter-alpha-trypsi
452	34	40.0	412	2	T14668	hypothetical prote	525	912	2	H90567	hypothetical prote
453	34	40.0	414	2	S30397	CD44 protein - mou	526	916	2	D83093	secretion protein
454	34	40.0	425	2	S54010	1-aminocyclopropan	527	918	2	S44769	hypothetical prote
455	34	40.0	433	1	UKMS	u-plasminogen acti	528	925	2	T22388	hypothetical prote
456	34	40.0	437	1	A45569	adenosylhomocyste	529	955	2	E84022	hypothetical prote
457	34	40.0	437	2	T32318	monoclonal antibod	530	984	2	AE0290	insecticidal toxin
458	34	40.0	444	2	FC4436	intracellular alka	531	990	1	G46335	env polyprotein pr
459	34	40.0	444	2	B83891	suppressor protein	532	1039	2	G83748	alpha-mannosidase
460	34	40.0	444	2	S48696	conserved hypothet	533	1042	1	GBECE	beta-galactosidase
461	34	40.0	450	2	B82973	hypothetical prote	534	1042	2	E85968	evolved beta-D-gal
462	34	40.0	454	2	T47578	hypothetical prote	535	1042	2	F91123	evolved beta-D-gal
463	34	40.0	456	2	AH0842	probable glycopori	536	1042	2	T26644	hypothetical prote
464	34	40.0	459	2	AC0075	probable membrane	537	1050	2	H90316	hypothetical prote
465	34	40.0	466	2	D84807	hypothetical prote	538	1050	2	A89769	hypothetical prote
466	34	40.0	468	2	D69716	involved in spore	539	1086	2	T17628	endo-1,4-beta-xyla
467	34	40.0	472	2	T24316	hypothetical prote	540	1103	2	T42022	probable chitin sy



541	34	40.0	1119	2	T16720	hypothetical prote	614	33	38.8	178	2	S50637	hypothetical prote
542	34	40.0	1205	2	T27053	hypothetical prote	615	33	38.8	179	2	S28256	NADH2 dehydrogenas
543	34	40.0	1232	2	T31436	cellulase (EC 3.2.	616	33	38.8	184	2	T22007	hypothetical prote
544	34	40.0	1233	2	T40059	chromosome segrega	617	33	38.8	200	2	F83008	hypothetical prote
545	34	40.0	1241	2	S01827	period clock prote	618	33	38.8	204	2	B75539	probable phosphogl
546	34	40.0	1286	1	H36845	DNA-directed RNA p	619	33	38.8	204	2	G87575	hypothetical prote
547	34	40.0	1286	2	T37366	RNA polymerase sub	620	33	38.8	204	2	F83067	hypothetical prote
548	34	40.0	1286	2	T28521	DNA-directed RNA p	621	33	38.8	205	2	T11657	rho GTP dissociati
549	34	40.0	1287	2	A72161	MGR protein - vari	622	33	38.8	209	2	T29765	hypothetical prote
550	34	40.0	1287	1	RNV247	DNA-directed RNA p	623	33	38.8	211	2	A72426	hypothetical prote
551	34	40.0	1310	2	AE1749	glycosidase homolo	624	33	38.8	215	2	T39583	hypothetical prote
552	34	40.0	1310	2	S78457	exonuclease II - f	625	33	38.8	216	2	AH2547	hypothetical prote
553	34	40.0	1354	2	AG0538	Rhs-family protein	626	33	38.8	220	2	T18792	hypothetical prote
554	34	40.0	1379	2	T45119	FIM protein [impor	627	33	38.8	221	2	S24322	glutathione transf
555	34	40.0	1533	2	F71274	hypothetical prote	628	33	38.8	221	2	A54858	glutathione transf
556	34	40.0	1609	2	S44821	F44E2.4 protein -	629	33	38.8	224	2	T16705	hypothetical prote
557	34	40.0	2197	2	B71600	variant-specific s	630	33	38.8	225	2	D84525	hypothetical prote
558	34	40.0	2251	2	T24490	hypothetical prote	631	33	38.8	225	2	D84526	hypothetical prote
559	34	40.0	2464	1	QRMSP1	microtubule-associ	632	33	38.8	229	1	G64371	conserved hypothet
560	34	40.0	2910	2	T28156	DNA-directed RNA p	633	33	38.8	230	2	B95353	protein [imported
561	34	40.0	3084	1	MMMSA	laminin alpha-1 ch	634	33	38.8	232	2	T25745	hypothetical prote
562	34	40.0	4924	2	T50176	probable peptide s	635	33	38.8	243	2	S11226	MyD88 protein - mo
563	33.5	39.4	60	2	A11991	hypothetical prote	636	33	38.8	244	2	F82510	probable transcrip
564	33.5	39.4	138	2	G82015	probable DnaK supp	637	33	38.8	248	2	AC0712	probable outer mem
565	33.5	39.4	206	2	G81071	conserved hypothet	638	33	38.8	251	2	T34168	hypothetical prote
566	33.5	39.4	226	2	T02467	probable transcrip	639	33	38.8	252	2	H84423	hypothetical prote
567	33.5	39.4	238	2	A72374	hypothetical prote	640	33	38.8	255	2	T13437	acid phosphatase h
568	33.5	39.4	260	2	T08503	trbF protein - Ent	641	33	38.8	258	2	A75618	probable chromosom
569	33.5	39.4	266	2	T20142	hypothetical prote	642	33	38.8	258	2	C88382	protein W06E11.4 f
570	33.5	39.4	319	2	AE2294	hypothetical prote	643	33	38.8	270	2	I51543	MHC class II beta-
571	33.5	39.4	341	2	T37502	hypothetical prote	644	33	38.8	271	2	D69790	probable halide pe
572	33.5	39.4	370	1	XNVKUD	UDPglucose-hexose-	645	33	38.8	271	2	D84586	hypothetical prote
573	33.5	39.4	388	2	F10208	oligogalacturonide	646	33	38.8	277	2	A32700	mannose 6-phosphat
574	33.5	39.4	465	2	F89857	conserved hypothet	647	33	38.8	278	1	A40399	mannose 6-phosphat
575	33.5	39.4	564	2	B83654	DNA polymerase III	648	33	38.8	279	2	A27068	mannose 6-phosphat
576	33.5	39.4	571	2	F70040	sulfite reductase	649	33	38.8	279	2	AD2516	hypothetical prote
577	33.5	39.4	600	2	A82043	inner membrane cop	650	33	38.8	281	2	T50208	yeast ktl12 protei
578	33.5	39.4	681	2	T01469	hypothetical prote	651	33	38.8	282	2	C97271	probable xylanase/
579	33.5	39.4	691	2	T33637	hypothetical prote	652	33	38.8	282	2	B81264	probable DNA ligas
580	33.5	39.4	708	2	T00205	hypothetical prote	653	33	38.8	288	2	C83487	probable transcrip
581	33.5	39.4	728	2	S57142	hypothetical prote	654	33	38.8	291	2	S66770	probable membrane
582	33.5	39.4	860	2	B75589	GGDEF family prote	655	33	38.8	295	2	F86541	HAD type hydrolase
583	33.5	39.4	1060	2	E83547	proline dehydrogen	656	33	38.8	295	2	F72082	had homolog hydrol
584	33.5	39.4	1145	2	S13643	PRP22 protein - ye	657	33	38.8	295	2	S61644	hypothetical prote
585	33.5	39.4	1148	2	H90175	NADH dehydrogenase	658	33	38.8	297	2	C82772	phage-related base
586	33.5	39.4	1194	2	T03818	apoptotic proteina	659	33	38.8	302	2	B49941	devR protein - Myx
587	33.5	39.4	1232	2	A55478	neuronal apoptosis	660	33	38.8	305	2	B72365	clostripain-relate
588	33	38.8	30	1	OE0N2K	beta-endorphin II	661	33	38.8	306	2	D69753	viomycin phosphotr
589	33	38.8	64	2	T16998	S-like ribonucleas	662	33	38.8	308	2	B84362	branched-chain ami
590	33	38.8	70	2	S54439	hemSTUV operon pro	663	33	38.8	310	2	F97112	methionyl-tRNA for
591	33	38.8	80	2	AB33067	hypothetical prote	664	33	38.8	312	2	T33160	hypothetical prote
592	33	38.8	86	2	B69895	spore coat protein	665	33	38.8	313	2	C86792	hypothetical prote
593	33	38.8	100	2	G86174	hypothetical prote	666	33	38.8	320	2	F96570	unknown protein, 8
594	33	38.8	102	2	A95384	protein [imported	667	33	38.8	323	2	S72473	type II site-speci
595	33	38.8	102	2	T19796	hypothetical prote	668	33	38.8	328	2	GH7529	D-3-phosphoglycera
596	33	38.8	124	2	D81190	hypothetical prote	669	33	38.8	328	2	AH2748	sugar kinase, ribo
597	33	38.8	125	2	S42425	aspartate kinase (	670	33	38.8	329	2	D97230	probable riboflavi
598	33	38.8	126	2	D72265	hypothetical prote	671	33	38.8	331	2	T44902	probable riboflavi
599	33	38.8	128	2	S63983	bile acid-binding	672	33	38.8	331	2	A70884	probable ribF prot
600	33	38.8	128	2	A32675	gastrotropin - pig	673	33	38.8	331	2	T15458	hypothetical prote
601	33	38.8	128	2	A54797	ileal lipid-bindin	674	33	38.8	339	2	S77404	cell division inh
602	33	38.8	128	2	JC1419	gastrotropin - rat	675	33	38.8	340	2	T32646	hypothetical prote
603	33	38.8	129	2	T22430	hypothetical prote	676	33	38.8	343	2	B86668	oligopeptide ABC t
604	33	38.8	131	2	B75433	hypothetical prote	677	33	38.8	344	2	T01572	sterol 24-C-methyl
605	33	38.8	134	2	T23217	hypothetical prote	678	33	38.8	344	2	T40138	sterol 24-C-methyl
606	33	38.8	135	2	AB2373	mannose-6-phosphat	679	33	38.8	348	2	F96997	uncharacterized co
607	33	38.8	143	2	T45922	probable C2H2-type	680	33	38.8	350	2	E75047	acyl carrier prote
608	33	38.8	142	2	AC2627	hypothetical prote	681	33	38.8	350	2	F71113	probable acyl carr
609	33	38.8	148	2	S23414	hypothetical prote	682	33	38.8	350	2	G84715	hypothetical prote
610	33	38.8	156	2	JC5752	iron-uptake regula	683	33	38.8	356	2	B64009	hypothetical prote
611	33	38.8	157	2	AF1132	hypothetical prote	684	33	38.8	358	2	T20814	hypothetical prote
612	33	38.8	171	2	A29795	circumsporozoite p	685	33	38.8	359	2	A12368	hypothetical prote
613	33	38.8	175	2	B86425	probable diadenosi	686	33	38.8	363	2	JC2543	angiotensin II rec

687 33 38.8 363 2 T06795 probable sterol 24  
688 33 38.8 363 2 G82556 conserved hypochet  
689 33 38.8 364 2 P96603 hypothetical prote  
690 33 38.8 378 2 T51647 myb-related trans  
691 33 38.8 379 2 S29978 hypD protein - Alc  
692 33 38.8 381 2 T36150 probable 4-hydroxy  
693 33 38.8 384 2 G96689 probable fructokin  
694 33 38.8 385 2 S32877 hypD protein - Rhi  
695 33 38.8 388 2 A39756 circumsporozoite p  
696 33 38.8 389 2 C86548 hypothetical prote  
697 33 38.8 389 2 B72076 hypothetical prote  
698 33 38.8 390 2 E97478 cycH protein [limp  
699 33 38.8 390 2 AE2696 cycH protein [limp  
700 33 38.8 393 2 T36842 probable secreted  
701 33 38.8 393 2 T49257 protein kinase-lik  
702 33 38.8 395 2 T24578 hypothetical prote  
703 33 38.8 398 2 C84780 hypothetical prote  
704 33 38.8 400 2 AF2749 isomerase/lactoniz  
705 33 38.8 400 2 E97530 dgoa protein [limp  
706 33 38.8 405 1 C8QPR ubiquinol-cytochro  
707 33 38.8 405 2 S05428 circumsporozoite p  
708 33 38.8 405 2 T21433 hypothetical prote  
709 33 38.8 407 2 T19155 hypothetical prote  
710 33 38.8 409 2 T06767 probable transcrip  
711 33 38.8 410 2 B55523 aromatic dioxygena  
712 33 38.8 410 2 S74705 hypothetical prote  
713 33 38.8 411 2 AE2993 glycosyltransferas  
714 33 38.8 411 2 E98290 hypothetical 42.6K  
715 33 38.8 412 1 OZZQAF circumsporozoite p  
716 33 38.8 417 2 S58193 adenosylhomocyste  
717 33 38.8 421 2 S42422 aspartate kinase (  
718 33 38.8 421 2 F70794 probable ask prote  
719 33 38.8 422 2 E84554 probable phosphose  
720 33 38.8 422 2 T40003 hypothetical prote  
721 33 38.8 424 2 A54533 circumsporozoite p  
722 33 38.8 427 2 E83801 GTP-binding protei  
723 33 38.8 428 2 S15662 farnesyltransferas  
724 33 38.8 430 2 S04104 NADH2 dehydrogenas  
725 33 38.8 430 2 T04668 phosphoserine tran  
726 33 38.8 441 2 S37881 hypothetical prote  
727 33 38.8 442 2 A54529 circumsporozoite p  
728 33 38.8 442 2 C90224 s-adenosyl-L-homoc  
729 33 38.8 443 2 T39538 alpha-amylase - Ae  
730 33 38.8 443 2 F71549 hypothetical prote  
731 33 38.8 444 2 E83759 alkaline phosphata  
732 33 38.8 452 2 T21435 hypothetical prote  
733 33 38.8 463 2 JE0193 NADH2 dehydrogenas  
734 33 38.8 463 2 H71462 probable fumarate  
735 33 38.8 467 2 JC6150 Gram-negative bact  
736 33 38.8 468 2 A37176 glutamate-ammonia  
737 33 38.8 476 2 T23213 hypothetical prote  
738 33 38.8 481 2 T15440 hypothetical prote  
739 33 38.8 488 2 S64140 hypothetical prote  
740 33 38.8 489 2 H89979 hypothetical prote  
741 33 38.8 499 2 S52422 chitinase (BC 3.2.  
742 33 38.8 499 2 S04856 chitinase (BC 3.2.  
743 33 38.8 500 2 T24901 hypothetical prote  
744 33 38.8 501 2 A55749 spliceosome-associ  
745 33 38.8 501 2 T49474 hypothetical prote  
746 33 38.8 507 1 RNNZCV polymerase-associ  
747 33 38.8 508 2 G95421 probable ABC trans  
748 33 38.8 510 2 T21430 hypothetical prote  
749 33 38.8 518 2 AI3540 probable binding p  
750 33 38.8 519 2 B45697 immediate-early pr  
751 33 38.8 523 2 C84753 hypothetical prote  
752 33 38.8 528 2 T24160 hypothetical prote  
753 33 38.8 529 2 JC5533 scavenger receptor  
754 33 38.8 529 2 S62194 hypothetical prote  
755 33 38.8 529 2 F82983 conserved hypochet  
756 33 38.8 530 2 S38092 hypothetical prote  
757 33 38.8 531 2 T08760 hypothetical prote  
758 33 38.8 544 2 A47726 dis1-suppressing p  
759 33 38.8 544 2 T33712 hypothetical prote

760 33 38.8 547 2 T00977  
761 33 38.8 549 2 S61706  
762 33 38.8 551 2 S66740  
763 33 38.8 564 2 A87384  
764 33 38.8 584 2 S27500  
765 33 38.8 587 2 T49357  
766 33 38.8 587 2 T00316  
767 33 38.8 587 2 S67555  
768 33 38.8 591 2 S73708  
769 33 38.8 595 2 T05087  
770 33 38.8 598 2 S34203  
771 33 38.8 598 2 AG0890  
772 33 38.8 605 2 T43974  
773 33 38.8 610 1 GVBPT4  
774 33 38.8 610 2 T44161  
775 33 38.8 613 2 T51938  
776 33 38.8 614 2 T22289  
777 33 38.8 615 1 S06977  
778 33 38.8 617 2 G64972 yegA protein precu  
779 33 38.8 617 2 B85933 suppressor of ompF  
780 33 38.8 617 2 G90987 suppressor of ompF  
781 33 38.8 619 2 AI0385 1-deoxy-D-xylulose  
782 33 38.8 620 2 AF0554 1-deoxy-D-xylulose  
783 33 38.8 620 2 D64771 dxs protein - Esch  
784 33 38.8 620 2 F85538 1-deoxy-D-xylulose  
785 33 38.8 620 2 B90688 1-deoxy-D-xylulose  
786 33 38.8 620 2 S56790 probable membrane  
787 33 38.8 623 2 C84694 probable protein k  
788 33 38.8 623 2 T09306 ERFR2 protein - hu  
789 33 38.8 623 2 T21112 G protein-coupled  
790 33 38.8 646 2 T51937 probable inositol-A  
791 33 38.8 647 2 E82579 acetyl coenzyme A  
792 33 38.8 647 2 C87693 acetyl-CoA synthet  
793 33 38.8 664 2 B86539 glycosyl hydrolase  
794 33 38.8 664 2 B72084 glycosyl hydrolase  
795 33 38.8 666 2 D82339 acetyl-CoA synthas  
796 33 38.8 670 2 AI3281 acetate-CoA ligase  
797 33 38.8 680 2 T51076 heterokaryon inco  
798 33 38.8 681 2 AD3045 conserved hypochet  
799 33 38.8 683 2 I39463 gene anonymous pro  
800 33 38.8 686 2 H98240 hypothetical prote  
801 33 38.8 717 2 B32838 DNA-directed RNA p  
802 33 38.8 719 2 S61046 ARP1 protein - yea  
803 33 38.8 723 2 T14765 hypothetical prote  
804 33 38.8 724 2 T23322 hypothetical prote  
805 33 38.8 725 2 S17444 phenylalanine ammo  
806 33 38.8 734 2 B83405 probable hydroxyli  
807 33 38.8 736 2 T40510 beta transducin -  
808 33 38.8 737 2 T46101 ABC transporter-li  
809 33 38.8 743 2 T09173 EH domain protein  
810 33 38.8 748 2 T02560 cellulose synthase  
811 33 38.8 755 2 T19118 acid phosphatase h  
812 33 38.8 759 2 AB2052 two-component hybr  
813 33 38.8 764 2 T05409 hypothetical prote  
814 33 38.8 772 2 T08226 hepatocyte growth  
815 33 38.8 775 2 I49759 tricorin proteinase  
816 33 38.8 784 2 B90442 replication licens  
817 33 38.8 795 1 S22804 hypothetical prote  
818 33 38.8 812 2 AC2349 hypothetical prote  
819 33 38.8 820 2 AD2357 hypothetical prote  
820 33 38.8 835 2 E72305 hypothetical prote  
821 33 38.8 835 2 A88238 protein T13H5.4 [i  
822 33 38.8 842 2 B96641 hypothetical prote  
823 33 38.8 842 2 T04208 probable anthranil  
824 33 38.8 857 2 G84844 protein F23F12.8 [i  
825 33 38.8 884 2 A90958 probable phase rep  
826 33 38.8 898 2 T42131 probable toxR-regu  
827 33 38.8 905 2 S29329 hypothetical prote  
828 33 38.8 908 2 T50695 secA protein [limp  
829 33 38.8 911 2 JN0921 transferin-bindin  
830 33 38.8 913 2 T35718 hypothetical prote  
831 33 38.8 918 2 T34057 hypothetical prote  
832 33 38.8 924 2 T50996 probable AMP deami

833	33	38.8	928	2	S40745	hypothetical prote	906	32.5	38.2	463	2	T31570	hypothetical prote
834	33	38.8	939	2	H81686	valyl-tRNA synthet	907	32.5	38.2	481	2	H69593	6-phospho-beta-glu
835	33	38.8	940	2	B85806	hypothetical prote	908	32.5	38.2	491	2	A59297	pre-B cell enhanci
836	33	38.8	947	2	T20512	hypothetical prote	909	32.5	38.2	493	2	S50443	probable phosphodi
837	33	38.8	958	2	E83701	alpha-amylase G-6	910	32.5	38.2	501	2	S44258	sucrose-6-phosphat
838	33	38.8	969	2	T38478	RhoGAP/Lim domain	911	32.5	38.2	528	2	AG1937	glucose-6-phosphat
839	33	38.8	970	2	A41944	mitotic control pr	912	32.5	38.2	541	2	T43862	PAD flavoprotein o
840	33	38.8	997	2	S63084	probable signaling	913	32.5	38.2	545	2	A49229	merozoite protein o
841	33	38.8	1010	2	G88554	protein F54C8.3 [i	914	32.5	38.2	562	2	T49386	hypothetical prote
842	33	38.8	1018	2	T30853	antigenic heat-sta	915	32.5	38.2	577	2	T20703	hypothetical prote
843	33	38.8	1026	2	C97783	cell surface antig	916	32.5	38.2	587	2	A34084	extracellular prot
844	33	38.8	1087	2	T22847	hypothetical prote	917	32.5	38.2	590	2	I56526	interleukin 1 rece
845	33	38.8	1091	2	T18666	lodestar maternal-	918	32.5	38.2	709	2	B82580	alanyl dipeptidyl
846	33	38.8	1095	1	A31225	phospholipase C (E	919	32.5	38.2	763	2	T22843	hypothetical prote
847	33	38.8	1116	2	B88612	protein Y76A2A.2 [	920	32.5	38.2	827	2	S59121	SOX6 protein - mou
848	33	38.8	1125	1	F70177	transcription-repa	921	32.5	38.2	848	2	S24981	adenylate cyclase
849	33	38.8	1131	2	A23944	chitin synthase (E	922	32.5	38.2	889	2	H84506	probable retroelem
850	33	38.8	1133	1	GNVUSR	M polyprotein prec	923	32.5	38.2	932	2	T28820	hypothetical prote
851	33	38.8	1133	1	A43964	M polyprotein prec	924	32.5	38.2	968	2	T29532	hypothetical prote
852	33	38.8	1133	2	S12597	M polyprotein prec	925	32.5	38.2	1087	2	I51552	platelet-derived g
853	33	38.8	1134	1	A43960	M polyprotein prec	926	32.5	38.2	1164	2	T01871	RNA-directed DNA p
854	33	38.8	1134	1	GNVU22	M polyprotein prec	927	32.5	38.2	1329	2	D87226	conserved hypothet
855	33	38.8	1148	2	F69685	pyruvate carboxyla	928	32.5	38.2	1577	2	T30858	glucosyltransferas
856	33	38.8	1186	2	T12737	tail protein - Met	929	32.5	38.2	3587	2	I40486	surfactin syntheta
857	33	38.8	1188	2	C71231	hypothetical prote	930	32.5	38.2	4450	2	JX0340	gramicidin S synth
858	33	38.8	1213	2	T19835	hypothetical prote	931	32.5	38.2	4452	1	YGBSG2	gramicidin S synth
859	33	38.8	1221	2	A49457	fibulin-2 precursor	932	32	37.6	65	1	RSBPX8	excisionase - phag
860	33	38.8	1223	2	S62011	PHO85 protein - ye	933	32	37.6	81	2	AF2998	hypothetical prote
861	33	38.8	1237	2	A54080	protein-tyrosine-p	934	32	37.6	109	2	T51864	probable heat-shoc
862	33	38.8	1238	1	JC5673	copper-transportin	935	32	37.6	113	2	J00810	transcription repr
863	33	38.8	1244	2	T19615	hypothetical prote	936	32	37.6	114	2	S76565	hypothetical prote
864	33	38.8	1247	2	E71616	hypothetical prote	937	32	37.6	115	2	G72674	hypothetical prote
865	33	38.8	1250	2	T22845	hypothetical prote	938	32	37.6	118	2	AH3186	conserved hypothet
866	33	38.8	1272	2	H82926	conserved hypothet	939	32	37.6	124	2	D75606	hypothetical prote
867	33	38.8	1308	2	B75198	DNA helicase relat	940	32	37.6	125	2	C70977	hypothetical prote
868	33	38.8	1317	2	S77517	RNA-directed RNA p	941	32	37.6	126	2	B82349	hypothetical prote
869	33	38.8	1350	2	AF2005	RNA polymerase bet	942	32	37.6	130	2	AF1885	hypothetical prote
870	33	38.8	1394	2	S66876	ATP-dependent tran	943	32	37.6	130	2	AB2910	conserved hypothet
871	33	38.8	1401	2	T17452	Werner syndrome pr	944	32	37.6	130	2	H97684	BH3119 conserved h
872	33	38.8	1401	2	T30247	Werner syndrome pr	945	32	37.6	140	2	T03347	gene e20 protein -
873	33	38.8	1482	2	I49704	Glutamate receptor	946	32	37.6	142	2	S32960	hypothetical prote
874	33	38.8	1482	2	B43274	N-methyl-D-asparta	947	32	37.6	147	2	H72570	probable ribosomal
875	33	38.8	1484	2	S52086	N-methyl-D-asparta	948	32	37.6	148	2	B97017	probable transcrip
876	33	38.8	1489	2	G71406	probable retroviru	949	32	37.6	151	2	T10196	hypothetical prote
877	33	38.8	1590	2	B86398	protein T7N9.24 [i	950	32	37.6	152	2	T17988	hypothetical prote
878	33	38.8	1663	2	T42092	s-afadin - rat	951	32	37.6	152	2	AG1774	hypothetical prote
879	33	38.8	1816	2	F83901	hypothetical prote	952	32	37.6	154	2	B87132	very hypothetical
880	33	38.8	1829	2	T41751	l-afadin - rat	953	32	37.6	156	2	B44048	hypothetical 18.1K
881	33	38.8	1881	2	H95076	zinc metalloprotei	954	32	37.6	163	2	S68520	Pin1 protein - hum
882	33	38.8	2094	2	S33124	tptr protein - huma	955	32	37.6	163	2	B96842	hypothetical prote
883	33	38.8	3026	2	T28431	variant surface pr	956	32	37.6	165	2	JC7136	peptidylprolyl iso
884	33	38.8	5170	2	T15348	hypothetical prote	957	32	37.6	167	2	AF3206	acetyltransferase
885	33	38.8	7829	2	T15789	hypothetical prote	958	32	37.6	167	2	D89886	conserved hypothet
886	32.5	38.2	97	2	JQ1490	HMG-containing pro	959	32	37.6	168	2	AF2820	conserved hypothet
887	32.5	38.2	109	2	S57400	hypothetical prote	960	32	37.6	168	2	AF2820	VPS23-like phospho
888	32.5	38.2	138	2	F81243	Diak suppressor pr	961	32	37.6	178	2	T36549	hypothetical prote
889	32.5	38.2	152	2	D90806	curlin major subun	962	32	37.6	178	2	T32796	hypothetical prote
890	32.5	38.2	152	2	H85665	hypothetical prote	963	32	37.6	180	2	F97402	hypothetical prote
891	32.5	38.2	212	2	AH1944	phosphoglycerate m	964	32	37.6	181	2	F81892	hypothetical prote
892	32.5	38.2	222	2	F83350	hypothetical prote	965	32	37.6	181	2	D81132	hypothetical prote
893	32.5	38.2	303	2	AH2016	hypothetical prote	966	32	37.6	182	2	D75474	conserved hypothet
894	32.5	38.2	307	2	D95148	aspartate carbamoy	967	32	37.6	187	2	G71312	probable translati
895	32.5	38.2	333	2	G95067	hypothetical prote	968	32	37.6	192	2	T26544	hypothetical prote
896	32.5	38.2	333	2	F97935	hypothetical prote	969	32	37.6	195	1	MFIV2C	matrix protein M2
897	32.5	38.2	349	2	F83275	conserved hypothet	970	32	37.6	195	1	MFIV2C	matrix protein M2
898	32.5	38.2	365	2	T27460	hypothetical prote	971	32	37.6	195	1	MFIVB2	matrix protein M2
899	32.5	38.2	377	2	T40024	probable cytochrom	972	32	37.6	197	2	F86798	prophage pi3 prote
900	32.5	38.2	378	2	S50040	granulocyte-macrop	973	32	37.6	198	2	D88098	protein T0604.5 [i
901	32.5	38.2	385	2	E70154	DNA-directed DNA p	974	32	37.6	198	2	T32025	hypothetical prote
902	32.5	38.2	401	2	B71115	acetate kinase [im	975	32	37.6	199	2	AD1230	phosphoglycerate m
903	32.5	38.2	438	1	XXMSN	phosphatidylcholin	976	32	37.6	199	2	AG1583	weakly phosphoglyc
904	32.5	38.2	438	2	B71963	probable outer mem	977	32	37.6	200	1	A26694	smooth muscle prot
905	32.5	38.2	450	2	A64546	hypothetical prote	978	32	37.6	204	2	D98206	hypothetical prote

979 32 37.6 204 2 AE3080 hypothetical prote  
 980 32 37.6 204 2 C97084 hypothetical prote  
 981 32 37.6 205 2 T15450 hypothetical prote  
 982 32 37.6 207 2 A75475 probable acetyltra  
 983 32 37.6 210 2 S67771 endoplasmic reticu  
 984 32 37.6 211 2 S15167 fixJ protein - Azo  
 985 32 37.6 217 2 T37832 hypothetical prote  
 986 32 37.6 217 2 E72047 conserved hypotet  
 987 32 37.6 219 2 A64340 hypothetical prote  
 988 32 37.6 220 2 T20880 hypothetical prote  
 989 32 37.6 221 2 JC4740 chaparonin grpe -  
 990 32 37.6 227 2 AB2535 hypothetical prote  
 991 32 37.6 228 2 S20416 ASP-56 protein - p  
 992 32 37.6 229 2 S58210 bradyzoite-specifi  
 993 32 37.6 232 2 B64442 hypothetical prote  
 994 32 37.6 234 2 AF0857 conserved hypotet  
 995 32 37.6 236 2 AH3247 conserved hypotet  
 996 32 37.6 239 2 C64188 arginine binding p  
 997 32 37.6 241 2 A40975 caffeoyl-CoA O-met  
 998 32 37.6 244 2 T15223 hypothetical prote  
 999 32 37.6 249 2 B69693 ribonuclease III (  
 1000 32 37.6 249 2 A84751 probable SF16 prot

ALIGNMENTS

RESULT 1

A38175  
 glucosyltransferase precursor - Streptococcus sobrinus  
 C;Species: Streptococcus sobrinus  
 C;Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Mar-2004  
 C;Accession: A38175  
 R;Abo, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kato, K.; Kagawa, H.  
 J. Bacteriol. 173, 989-996, 1991  
 A;Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus  
 A;Reference number: A38175; MUID:91123227; PMID:1704006  
 A;Accession: A38175  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-1592 <ABO>  
 A;Cross-references: UNIPARC:UPI000012BCB2; GB:D90213; NID:g217032; PIDN:BAA14241.1; PID:  
 F;1093-1114/Domain: cpl repeat homology <CP1>  
 F;1222-1241/Domain: cpl repeat homology <CP2>  
 F;1287-1306/Domain: cpl repeat homology <CP3>  
 F;1330-1351/Domain: cpl repeat homology <CP4>  
 F;1352-1371/Domain: cpl repeat homology <CP5>  
 F;1402-1420/Domain: cpl repeat homology <CP6>  
 F;1465-1484/Domain: cpl repeat homology <CP7>  
 F;1513-1532/Domain: cpl repeat homology <CP8>

Query Match 100.0%; Score 85; DB 2; Length 1592;  
 Best Local Similarity 100.0%; Pred. No. 9.5e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWNGSEKPYDDHL 14  
 |||||  
 DB 337 QWNGSEKPYDDHL 350

RESULT 2

JT0345  
 dextranase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)  
 N;Alternate names: sucrose 6-glucosyltransferase  
 C;Species: Streptococcus mutans  
 C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
 C;Accession: JT0345; C33135  
 R;Ueda, S.; Shiroza, T.; Kuramitsu, H.K.  
 Gene 69, 101-109, 1988  
 A;Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.  
 A;Reference number: JT0345; MUID:89137980; PMID:2976010  
 A;Accession: JT0345  
 A;Molecule type: DNA

A;Residues: 1-1375 <UED>  
 A;Cross-references: UNIPROT:P13470; UNIPARC:UPI0000155515  
 A;Experimental source: GS-5  
 R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
 J. Bacteriol. 169, 4263-4270, 1987  
 A;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
 A;Reference number: A33135; MUID:87308013; PMID:3040685  
 A;Accession: C33135  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-349 <SHI>  
 A;Cross-references: UNIPARC:UPI000014E25D; GB:M17361  
 C;Genetics:  
 A;Gene: gtfC  
 C;Function:  
 A;Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
 C;Keywords: duplication; glycosyltransferase; hexosyltransferase  
 F;1-34/Domain: signal sequence #status predicted <SIG>  
 F;35-1375/Product: glucosyltransferase #status predicted <MAT>  
 F;1126-1145/Domain: cpl repeat homology <CP1>  
 F;1253-1272/Domain: cpl repeat homology <CP2>  
 F;1318-1337/Domain: cpl repeat homology <CP3>

Query Match 78.8%; Score 67; DB 2; Length 1375;  
 Best Local Similarity 76.9%; Pred. No. 0.0084;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14  
 || : |||||  
 DB 370 WNSDSEKPFDDHL 382

RESULT 3

B33135  
 gtfB protein precursor - Streptococcus mutans  
 C;Species: Streptococcus mutans  
 C;Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 09-Jul-2004  
 C;Accession: B33135; A33128  
 R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
 J. Bacteriol. 169, 4263-4270, 1987  
 A;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
 A;Reference number: A33135; MUID:87308013; PMID:3040685  
 A;Accession: B33135  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-1475 <SHI>  
 A;Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:g153639; PIDN:  
 R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
 submitted to the Protein Sequence Database, September 1990  
 A;Reference number: A33128  
 A;Accession: A33128  
 A;Status: preliminary; not compared with conceptual translation  
 A;Molecule type: DNA  
 A;Residues: 1-171,173-641,'N',643-1475 <SH2>  
 A;Cross-references: UNIPARC:UPI000017AC5E  
 A;Experimental source: strain GS-5  
 F;1096-1115/Domain: cpl repeat homology <CP1>  
 F;1224-1243/Domain: cpl repeat homology <CP2>  
 F;1289-1308/Domain: cpl repeat homology <CP3>  
 F;1354-1373/Domain: cpl repeat homology <CP4>  
 F;1419-1438/Domain: cpl repeat homology <CP5>

Query Match 78.8%; Score 67; DB 2; Length 1475;  
 Best Local Similarity 76.9%; Pred. No. 0.0091;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14  
 || : |||||  
 DB 344 WNSDSEKPFDDHL 356

RESULT 4

T31098





R;Waterston, R.  
 submitted to the EMBL Data Library, April 1996  
 A;Description: The C. briggsae genome sequencing project.  
 A;Reference number: Z21010  
 A;Accession: T31345  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-4767 <WAT>  
 A;Cross-references: UNIPROT:Q17301; UNIPARC:UPI000011013B; EMBL:U56248; NID:g1293789; PI  
 C;Genetics:  
 A;Introns: 269/3; 341/3; 853/1; 920/2; 4452/3; 4534/3; 4592/3; 4654/3; 4670/3; 4707/2; 4  
 A;Note: G01D9.5  
 C;Keywords: carrier protein; phosphopantetheine; phosphoprotein  
 F;3472-3541/Domain: acyl carrier protein homology <ACPI>  
 F;4039-4427/Domain: acetate-CoA ligase homology <ACL>  
 F;4447-4514/Domain: acyl carrier protein homology <ACP2>  
 F;2210,3505/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 50.6%; Score 43; DB 2; Length 4767;  
 Best Local Similarity 70.0%; Pred. No. 3.5e+02;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GSESEKPYDDH 13  
 ||:||||:|  
 Db 1115 GENEKPFDDVH 1124

RESULT 14  
 RGEQWJ  
 Met regulon regulatory protein metJ - Escherichia coli (strain K-12)  
 C;Species: Escherichia coli  
 C;Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 12-Jul-2004  
 C;Accession: A22660; S40881; I54931; E65200  
 R;Saint-Giron, I.; Duchange, N.; Cohen, G.N.; Zakin, M.M.  
 J. Biol. Chem. 259, 14282-14285, 1984  
 A;Title: Structure and autoregulation of the metJ regulatory gene in Escherichia coli.  
 A;Reference number: A22660; MUID:85054884; PMID:6094549  
 A;Accession: A22660  
 A;Molecule type: DNA  
 A;Residues: 1-105 <SAI>  
 A;Cross-references: UNIPROT:P08338; UNIPARC:UPI0000165544; GB:M12869; NID:g146836; PIDN:  
 R;Plunkett III, G.; Burland, V.; Daniels, D.L.; Blattner, F.R.  
 Nucleic Acids Res. 21, 3391-3398, 1993  
 A;Title: Analysis of the Escherichia coli genome. III. DNA sequence of the region from 8  
 A;Reference number: S40802; MUID:93347969; PMID:8346018  
 A;Accession: S40881  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-105 <PLU>  
 A;Cross-references: UNIPARC:UPI0000165544; EMBL:L19201; NID:g304961; PIDN:AAB03070.1; PI  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1993  
 R;Collier, C.D.; Johnson, J.R.  
 J. Bacteriol. 172, 3918-3924, 1990  
 A;Title: The Escherichia coli K-12 metJ193 allele contains a point mutation which alters  
 duction of met regulon expression.  
 A;Reference number: I54931; MUID:90299817; PMID:2141834  
 A;Accession: I54931  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-105 <RRS>  
 A;Cross-references: UNIPARC:UPI0000165544; GB:M38202; NID:g146834; PIDN:AAA24162.1; PID:  
 R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
 Science 277, 1453-1462, 1997  
 A;Title: The complete genome sequence of Escherichia coli K-12.  
 A;Reference number: A64720; MUID:97428617; PMID:9278503  
 A;Accession: E65200  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-105 <BLAT>  
 A;Cross-references: UNIPARC:UPI0000165544; GB:AE000467; GB:U00096; NID:g1790356; PIDN:AP  
 A;Experimental source: strain K-12, substrain MGL55  
 C;Genetics:

A;Gene: metJ  
 A;Map position: 89 min  
 C;Function:  
 A;Description: when combined with methionine, represses the expression of the methionine  
 C;Superfamily: Met repressor  
 C;Keywords: DNA binding; transcription regulation  
 F;2-105/Product: Met regulon regulatory protein metJ #status predicted <MAT>

Query Match 49.4%; Score 42; DB 1; Length 105;  
 Best Local Similarity 46.2%; Pred. No. 7.5;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QWNGESEKPYDDH 13  
 :||:|  
 Db 3 EWSGEYISPYAEH 15

RESULT 15  
 AB0938  
 repressor of the methionine regulon [imported] - Salmonella enterica subsp. enterica ser  
 C;Species: Salmonella enterica subsp. enterica serovar Typhi  
 A;Note: this species has also been called Salmonella typhi  
 C;Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 12-Jul-2004  
 C;Accession: AB0938  
 R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 , S.; Moule, S.; O'Gaora, P.  
 Nature 413, 848-852, 2001  
 A;Authors: Parry, C.; Ouail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov  
 A;Reference number: AB0502; MUID:21534947; PMID:11677608  
 A;Accession: AB0938  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-105 <PAR>  
 A;Cross-references: UNIPARC:UPI000005A604; GB:AL513382; PIDN:CAD09524.1; PID:g16504641;  
 C;Genetics:  
 A;Gene: STY3770  
 C;Superfamily: Met repressor

Query Match 49.4%; Score 42; DB 2; Length 105;  
 Best Local Similarity 46.2%; Pred. No. 7.5;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QWNGESEKPYDDH 13  
 :||:|  
 Db 3 EWSGEYISPYAEH 15

RESULT 16  
 G86084  
 repressor of all met genes but metF [imported] - Escherichia coli (strain O157:H7, subst  
 C;Species: Escherichia coli  
 C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 12-Jul-2004  
 C;Accession: G86084  
 R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
 Nature 409, 529-533, 2001  
 A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
 A;Reference number: A85480; MUID:21074935; PMID:11206551  
 A;Accession: G86084  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-105 <STO>  
 A;Cross-references: UNIPROT:P08338; UNIPARC:UPI0000165544; GB:AE005174; NID:g12518853; P  
 A;Experimental source: strain O157:H7, substrain EDL933  
 C;Genetics:  
 A;Gene: metJ  
 C;Superfamily: Met repressor

Query Match 49.4%; Score 42; DB 2; Length 105;  
 Best Local Similarity 46.2%; Pred. No. 7.5;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

```

QY      1 QWNGSEKPYDDH 13
      :|:|:|:|:|
Db      3 EWSGEYISPYAEH 15

RESULT 17
C91237
repressor protein MetJ [imported] - Escherichia coli (strain O157:H7, substrain RIMD 050
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 12-Jul-2004
C;Accession: C91237
R;Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gaawata, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: C91237
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-105 <HAY>
A;Cross-references: UNIPROT:P08338; UNIPARC:UPI0000165544; GB:BA000007; PIDN:BA838290.1;
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4867
C;Superfamily: Met repressor

Query Match      49.4%; Score 42; DB 2; Length 105;
Best Local Similarity 46.2%; Pred. No. 7.5;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      1 QWNGSEKPYDDH 13
      :|:|:|:|:|
Db      3 EWSGEYISPYAEH 15

RESULT 18
A23081
regulatory protein metJ - Salmonella typhimurium
C;Species: Salmonella typhimurium
C;Date: 23-Aug-1987 #sequence_revision 23-Aug-1987 #text_change 12-Jul-2004
C;Accession: A23081
R;Urbanowski, M.L.; Stauffer, G.V.
Nucleic Acids Res. 13, 673-685, 1985
A;Title: Nucleotide sequence and biochemical characterization of the metJ gene from Salmo
A;Reference number: A23081; MUID:85215507; PMID:2987805
A;Accession: A23081
A;Molecule type: DNA
A;Residues: 1-105 <URB>
A;Cross-references: UNIPROT:P06203; UNIPARC:UPI00001622A8; GB:X01961; NID:g47780; PIDN:C
C;Genetics:
A;Gene: metJ
A;Map position: 87
C;Superfamily: Met repressor
C;Keywords: DNA binding; transcription regulation

Query Match      49.4%; Score 42; DB 2; Length 105;
Best Local Similarity 46.2%; Pred. No. 7.5;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      1 QWNGSEKPYDDH 13
      :|:|:|:|:|
Db      3 EWSGEYISPYAEH 15

RESULT 19
H83068
hypothetical protein PA4608 [imported] - Pseudomonas aeruginosa (strain PA01)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C;Accession: H83068
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,

```

```

.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83068
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-125 <STO>
A;Cross-references: UNIPROT:Q9HV11; UNIPARC:UPI00000C5D87; GB:AE004875; GB:AE004091; NID
A;Experimental source: strain PA01
C;Genetics:
A;Gene: PA4608

Query Match      49.4%; Score 42; DB 2; Length 125;
Best Local Similarity 38.5%; Pred. No. 9.1;
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY      2 WNGSEKPYDDHL 14
      |||:|:|:|
Db      49 WNGDPQRPPEARL 61

RESULT 20
B95201
3-hydroxy-3-methylglutaryl-CoA reductase [imported] - Streptococcus pneumoniae (strain T
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 09-Jul-2004
C;Accession: B95201
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple, I.
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: B95201
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-424 <KUR>
A;Cross-references: UNIPROT:Q97PB4; UNIPARC:UPI0000051976; GB:AE005672; PIDN:AAK75803.1;
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP1726

Query Match      49.4%; Score 42; DB 2; Length 424;
Best Local Similarity 53.8%; Pred. No. 35;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      2 WNGSEKPYDDHL 14
      |||:|:|:|
Db      5 WNGFSKKSQERL 17

RESULT 21
H98067
hydroxymethylglutaryl-CoA reductase (EC 1.1.1.88) [imported] - Streptococcus pneumoniae
C;Species: Streptococcus pneumoniae
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
C;Accession: H98067
R;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A;Reference number: A97872; MUID:21429245; PMID:11544234
A;Accession: H98067
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-424 <KUR>
A;Cross-references: UNIPROT:QBDN55; UNIPARC:UPI00000E3695; GB:AE007317; PIDN:AAL00373.1;
C;Genetics:
A;Gene: mvaa

```



Nature 335, 547-550, 1988

A;Cross-references: UNIPROT:O86806; UNIPARC:UPI000000DACA8; EMBL:AL031031; PIDN:CAA19847.  
 A;Experimental source: strain A3(2)  
 C;Genetics:  
 A;Gene: SC0DB:SC7C7.01

Query Match 48.2%; Score 41; DB 2; Length 225;  
 Best Local Similarity 77.8%; Pred. No. 26;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GESEKPYDD 12  
 |||||  
 Db 97 GESEAPYDE 105  
 |||||

RESULT 27  
 A87602  
 conserved hypothetical protein CC2849 [imported] - Caulobacter crescentus  
 C;Species: Caulobacter crescentus  
 C;Date: 20-Apr-2001 #sequence\_revision 20-Apr-2001 #text\_change 09-Jul-2004  
 C;Accession: A87602  
 R;Niernan, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.  
 B.; Laub, M.W.; DeBoy, R.T.; Dodson, R.O.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon  
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.  
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001  
 A;Title: Complete Genome Sequence of Caulobacter crescentus.  
 A;Reference number: A87249; MUID:21173698; PMID:11259647  
 A;Accession: A87602  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-389 <STO>  
 A;Cross-references: UNIPROT:Q9A412; UNIPARC:UPI000000C787C; GB:AE005673; NID:g13424461; E  
 C;Genetics:  
 A;Gene: CC2849

Query Match 48.2%; Score 41; DB 2; Length 389;  
 Best Local Similarity 66.7%; Pred. No. 47;  
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WNGESEKPY 10  
 |||||  
 Db 241 WNGEGPNPY 249  
 |||||

RESULT 28  
 E85135  
 hypothetical protein ATg12620 [imported] - Arabidopsis thaliana  
 C;Species: Arabidopsis thaliana (mouse-ear cress)  
 C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
 C;Accession: E85135  
 R;anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
 Nature 402, 769-777, 1999  
 A;Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
 A;Reference number: A85001; MUID:20083488; PMID:10617198  
 A;Accession: E85135  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-813 <STO>  
 A;Cross-references: UNIPROT:Q9SU24; UNIPARC:UPI000000A52D7; GB:NC\_001268; NID:g7267964; E  
 C;Genetics:  
 A;Gene: ATg12620  
 A;Map position: 4

Query Match 48.2%; Score 41; DB 2; Length 813;  
 Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QWNGSEKPYDD 12  
 |||||  
 Db 356 EWNGRKEEVD 367  
 |||||

RESULT 29  
 T32925  
 hypothetical protein K09H9.7 - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
 C;Accession: T32925  
 R;Madsen, C.; Graves, T.; Blair, T.  
 submitted to the EMBL Data Library, January 1998  
 A;Description: The sequence of C. elegans cosmid K09H9.  
 A;Reference number: Z21247  
 A;Accession: T32925  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-179 <MAD>  
 A;Cross-references: UNIPROT:O44993; UNIPARC:UPI0000077142; EMBL:AF043700; PIDN:AA897573.1  
 A;Experimental source: strain Bristol N2; clone K09H9  
 C;Genetics:  
 A;Gene: CESP:K09H9.7  
 A;Map position: 1  
 A;Introns: 9/1; 39/1; 83/1; 111/2; 165/2  
 C;Superfamily: Caenorhabditis elegans hypothetical protein K09H9.7

Query Match 47.1%; Score 40; DB 2; Length 179;  
 Best Local Similarity 50.8%; Pred. No. 29;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QWNGSEKPYDDHL 14  
 :||:|:|:|  
 Db 157 EWGGEAEKRIDQLL 170  
 :||:|:|:|

RESULT 30  
 A03308  
 dnaK-type molecular chaperone Hsp70Ab - fruit fly (Drosophila melanogaster) (fragments)  
 N;Alternate names: major heat shock 70K protein 1  
 C;Species: Drosophila melanogaster  
 C;Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 31-Dec-2004  
 C;Accession: A03308  
 R;Torok, I.; Karch, F.  
 Nucleic Acids Res. 8, 3105-3123, 1980  
 A;Title: Nucleotide sequences of heat shock activated genes in Drosophila melanogaster.  
 A;Reference number: A03308; MUID:81076551; PMID:6255408  
 A;Accession: A03308  
 A;Molecule type: DNA  
 A;Residues: 1-346 <TOR>  
 A;Cross-references: UNIPROT:P02825; UNIPARC:UPI0000177CE9  
 A;Note: this protein is coded by one of two genes at chromosome locus 87A7  
 C;Genetics:  
 A;Gene: FlyBase:Hsp70Ab  
 A;Cross-references: FlyBase:FBgn0013276  
 C;Function:  
 A;Description: involved in protein folding and assembling/disassembling of protein complex  
 C;Superfamily: bcr protein  
 C;Keywords: ATP; molecular chaperone

Query Match 47.1%; Score 40; DB 2; Length 346;  
 Best Local Similarity 46.2%; Pred. No. 61;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QWNGSEKPYDDH 13  
 |||||  
 Db 185 QWNPERQGGDEH 197  
 |||||

RESULT 31  
 AF3321  
 alpha-ketoglutarate permease [imported] - Brucella melitensis (strain 16M)  
 C;Species: Brucella melitensis  
 C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 09-Jul-2004  
 C;Accession: AF3321  
 R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova, N.  
 ; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letescu  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A;Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis  
 A;Reference number: AD3252; PMID:11756688

A;Accession: AF3321  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-457 <KUR>  
A;Cross-references: UNIPROT:Q8YI89; UNIPARC:UPI0000057D16; GB:AE008917; PIDN:AAL51737.1;  
A;Experimental source: strain 16M  
C;Genetics:  
A;Gene: BMEI0556  
A;Map position: 1  
C;Superfamily: citrate utilization determinant

Query Match 47.1%; Score 40; DB 2; Length 457;  
Best Local Similarity 63.6%; Pred. No. 83;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 NGESEKPYDDH 13  
||| :|||  
Db 22 NGEVEAPHDAH 32  
||| :|||

RESULT 32  
JC6170  
GATA-transcription factor - Penicillium chrysogenum  
C;Species: Penicillium chrysogenum  
C;Date: 11-Apr-1997 #sequence\_revision 09-May-1997 #text\_change 09-Jul-2004  
C;Accession: JC6170  
R;Haas, H.; Angermayr, K.; Stoeffler, G.  
Gene 184, 33-37, 1997  
A;Title: Molecular analysis of a Penicillium chrysogenum GATA factor encoding gene (sreP)  
A;Reference number: JC6170; MUID:97169143; PMID:9016950  
A;Accession: JC6170  
A;Molecule type: mRNA  
A;Residues: 1-532 <HAA>  
A;Cross-references: UNIPROT:Q92259; UNIPARC:UPI0000135F37; GB:U48414; NID:gl517915; PIDN:  
C;Comment: This factor is a member of GATA-binding protein family that recognizes a target  
C;Genetics:  
A;Gene: sreP  
A;Introns: 95/2; 263/1  
C;Keywords: transcription factor; zinc finger  
F;91-145/Domain: GATA-type zinc finger homology <GZF>  
F;235-288/Domain: GATA-type zinc finger homology <GZF2>

Query Match 47.1%; Score 40; DB 2; Length 532;  
Best Local Similarity 66.7%; Pred. No. 98;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 QWNGSEKPYD 9  
||| :|||  
Db 67 QWNGHNTP 75  
||| :|||

RESULT 33  
JC7537  
beta-N-acetylglucosaminidase - Alteromonas sp. (strain O-7)  
C;Species: Alteromonas sp. (strain O-7)  
C;Date: 30-Jun-2001 #sequence\_revision 30-Jun-2001 #text\_change 09-Jul-2004  
C;Accession: JC7537; PC7108  
R;Tsujibo, H.; Miyamoto, J.; Kondo, N.; Miyamoto, K.; Baba, N.; Inamori, Y.  
BioSci. Biotechnol. Biochem. 64, 2512-2516, 2000  
A;Title: Molecular cloning of the gene encoding an outer-membrane-associated beta-N-acetyl  
A;Reference number: JC7537; MUID: 21036923; PMID:11193430  
A;Accession: JC7537  
A;Molecule type: DNA  
A;Residues: 1-863 <TGU>  
A;Cross-references: UNIPROT:Q9FAC5; UNIPARC:UPI00000175B55; DDBJ:AB042622  
A;Experimental source: strain O-7  
A;Accession: PC7108  
A;Molecule type: protein  
A;Residues: 336-356; 464-471; 520-531; 548-557; 609-617; 716-726 <TS2>  
A;Cross-references: UNIPARC:UPI0000175B56; UNIPARC:UPI0000175B57; UNIPARC:UPI0000175B58;  
C;Comment: This enzyme, an outer-membrane-associated lipoprotein, which belongs to the f  
Glucosamine.  
C;Genetics:

A;Gene: GlcNAcase  
A;Superfamily: beta-hexosaminidase  
C;Keywords: lipoprotein

Query Match 47.1%; Score 40; DB 2; Length 863;  
Best Local Similarity 58.3%; Pred. No. 1.7e+02;  
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 3 NGESEKPYDDHL 14  
||| :|||  
Db 826 NGESWQPYTDQM 837  
||| :|||

RESULT 34  
S66704  
hypothetical protein YOL021c - yeast (Saccharomyces cerevisiae)  
N;Alternate names: hypothetical protein O2197  
C;Species: Saccharomyces cerevisiae  
C;Date: 12-Jul-1996 #sequence\_revision 12-Jul-1996 #text\_change 09-Jul-2004  
C;Accession: S66704  
R;Habbig, B.; Hattenhorst, U.; Hollenberg, C.P.; Ramezani Rad, M.  
submitted to the Protein Sequence Database, July 1996  
A;Reference number: S66703  
A;Accession: S66704  
A;Molecule type: DNA  
A;Residues: 1-1001 <HAB>  
A;Cross-references: UNIPROT:Q08162; UNIPARC:UPI0000052F2F; EMBL:Z74763; NID:gl419800; PI:  
A;Experimental source: strain S288C  
C;Genetics:  
A;Gene: SGD:D1S3  
A;Cross-references: SGD:S0005381; MIPS:YOL021C  
A;Map position: 15L  
C;Superfamily: mitotic control protein dis3

Query Match 47.1%; Score 40; DB 2; Length 1001;  
Best Local Similarity 63.6%; Pred. No. 2e+02;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QWNGSEKPYD 11  
||| :|||  
Db 179 QWYSEHLKPYD 189  
||| :|||

RESULT 35  
A44766  
defective chorion-1 fc125 protein precursor - fruit fly (Drosophila melanogaster)  
C;Species: Drosophila melanogaster  
C;Date: 03-Mar-1993 #sequence\_revision 03-Mar-1993 #text\_change 09-Jul-2004  
C;Accession: A44766  
R;Waring, G.L.; Hawley, R.J.; Schoenfeld, T.  
Dev. Biol. 142, 1-12, 1990  
A;Title: Multiple proteins are produced from the dec-1 eggshell gene in Drosophila by al.  
A;Reference number: A44766; MUID:91032553; PMID:1699826  
A;Accession: A44766  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1123 <WAR>  
A;Cross-references: UNIPROT:P18169; UNIPARC:UPI0000128F37; GB:M35887; NID:gl57181; PID:  
C;Genetics:  
A;Gene: FlyBase:dec-1  
A;Cross-references: FlyBase:FBgn0000427

Query Match 46.5%; Score 39.5; DB 2; Length 1123;  
Best Local Similarity 58.3%; Pred. No. 2.7e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

Qy 2 WNGSEKPYDDH 13  
||| :|||  
Db 1033 WN---QKPYDPH 1041  
||| :|||

RESULT 36  
E95909

hypothetical protein [imported] - Sinorhizobium meliloti (strain 1021) megaplasmid pSymB  
 C;Species: Sinorhizobium meliloti  
 C;Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 09-Jul-2004  
 C;Accession: E95909  
 R;Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan  
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001  
 A;Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-fixing endo  
 A;Reference number: A95842; MUID:21396508; PMID:11481431  
 A;Accession: E95909  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-304 <CUR>  
 A;Cross-references: UNIPROT:Q92W06; UNIPARC:UPI00000CB581; GB:AL591985; PIDN:CAC48941.1;  
 A;Experimental source: strain 1021, megaplasmid pSymB  
 R;Galibert, F.; Finan, T.M.; Long, S.R.; Publer, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
 pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
 L.; Hyman, R.W.; Jones, T.  
 Science 283, 668-672, 2001  
 A;Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
 hebaullt, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.  
 A;Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
 A;Reference number: A96039; MUID:21368234; PMID:11474104  
 A;Contents: annotation  
 C;Genetics:  
 A;Gene: SMB20562  
 A;Genome: plasmid

Query Match 45.9%; Score 39; DB 2; Length 304;  
 Best Local Similarity 46.2%; Pred. No. 78;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14  
 | : ||| ||| : :  
 Db 204 WDRSENPFYLDYM 216

RESULT 37  
 R6BY13  
 Ribosomal protein MRP13 precursor, mitochondrial - yeast (Saccharomyces cerevisiae)  
 N;Alternate names: protein G4618; protein YGR084c; ribosomal protein Yms-A  
 C;Species: Saccharomyces cerevisiae  
 C;Date: 30-Jun-1992 #sequence\_revision 19-Jul-1996 #text\_change 09-Jul-2004  
 C;Accession: S64379; A30237; S78035; S07831  
 R;Wedler, H.; Scharfe, M.; Wedler, E.; Wambutt, R.  
 submitted to the Protein Sequence Database, May 1996  
 A;Reference number: S64356  
 A;Accession: S64379  
 A;Molecule type: DNA  
 A;Residues: 1-339 <WED>  
 A;Cross-references: UNIPROT:P12686; UNIPARC:UPI0000053087; EMBL:Z72869; NID:g1323120; PD  
 R;Partaledis, J.A.; Mason, T.L.  
 Mol. Cell. Biol. 8, 3647-3660, 1988  
 A;Title: Structure and regulation of a nuclear gene in Saccharomyces cerevisiae that spe  
 A;Reference number: A30237; MUID:89127203; PMID:3065621  
 A;Accession: A30237  
 A;Molecule type: DNA  
 A;Residues: 28-113, 'S', 115-309, 'RIQRIIRAFESV', 323, 'TVGNLTSGSG', 336, 'RG', 339, 'IVQNTHRKYI  
 A;Cross-references: UNIPARC:UPI0000168CEF; EMBL:M22109; NID:g171970; PIDN:AAA34788.1; PI  
 A;Note: the authors translated the codon AGT for residues 4, 227, 236, and 282 as phe  
 R;Kitakawa, M.; Graack, H.R.; Grohmann, L.; Goldschmidt-Reisin, S.; Herfurth, E.; Wittma  
 Eur. J. Biochem. 245, 449-456, 1997  
 A;Title: Identification and characterization of the genes for mitochondrial ribosomal pr  
 A;Reference number: S78018; MUID:97296414; PMID:9151978  
 A;Accession: S78035  
 A;Molecule type: protein  
 A;Residues: 'X', 39-51 <KIT>  
 A;Cross-references: UNIPARC:UPI0000173B36  
 C;Genetics:  
 A;Gene: SGD:MRP13; MIPS:YGR084c  
 A;Cross-references: SGD:S0003316; MIPS:YGR084c  
 A;Map position: 7R  
 A;Genome: nuclear

C;Superfamily: Saccharomyces cerevisiae mitochondrial ribosomal protein MRP13  
 C;Keywords: mitochondrion; protein biosynthesis; ribosome  
 F;1-47/Domain: transit peptide (mitochondrion) #status predicted <INP>  
 F;48-339/Product: ribosomal protein MRP13 #status predicted <MAT>

Query Match 45.9%; Score 39; DB 1; Length 339;  
 Best Local Similarity 70.0%; Pred. No. 87;  
 Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 5 ESEKPYDDHL 14  
 || ||| |||  
 Db 310 ESSDPYDTHL 319

RESULT 38  
 A45545  
 major merozoite surface antigen - malaria parasite (Plasmodium falciparum) (fragments)  
 C;Species: Plasmodium falciparum  
 C;Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 31-Dec-2004  
 C;Accession: A45545  
 R;Blackman, M.J.; Ling, I.T.; Nicholls, S.C.; Holder, A.A.  
 Mol. Biochem. Parasitol. 49, 29-33, 1991  
 A;Title: Proteolytic processing of the Plasmodium falciparum merozoite surface protein-1  
 A;Reference number: A45545; MUID:92131048; PMID:1775158  
 A;Accession: A45545  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-400 <BLA>  
 A;Cross-references: UNIPROT:Q03999; UNIPARC:UPI000007B5E4  
 A;Note: sequence extracted from NCBI backbone (NCBIN:77612, NCBI:P:77621)  
 C;Superfamily: G surface protein  
 C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat

Query Match 45.9%; Score 39; DB 2; Length 400;  
 Best Local Similarity 77.8%; Pred. No. 11e+02;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 GESEKPYDD 12  
 ||||| : |||  
 Db 8 GESEEDYDD 16

RESULT 39  
 T31546  
 hypothetical protein Y47D3A.17 - Caenorhabditis elegans  
 C;Species: Caenorhabditis elegans  
 C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 31-Dec-2004  
 C;Accession: T31546  
 R;Matthews, L.  
 submitted to the EMBL Data Library, October 1999  
 A;Reference number: Z21043  
 A;Accession: T31546  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-449 <WIL>  
 A;Cross-references: UNIPROT:Q95Q23; UNIPROT:Q9U2C7; UNIPARC:UPI000017BC7E; EMBL:AL117202  
 A;Experimental source: clone Y47D3A  
 C;Genetics:  
 A;Gene: CESP:Y47D3A.17  
 A;Introns: 76/3; 113/3; 165/3; 271/3; 329/2; 370/2; 403/1  
 C;Superfamily: oxysterol-binding protein

Query Match 45.9%; Score 39; DB 2; Length 449;  
 Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 WNGSEKPY 9  
 ||| |||  
 Db 303 WNGNHEKP 310

RESULT 40  
 T28662

hypothetical protein P54 - Spiroplasma citri  
 C;Species: Spiroplasma citri  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C;Accession: T28662  
 R;Ye, F.; Melcher, U.; Rascoe, J.E.; Fletcher, J.  
 Biochem. Genet. 34, 269-286, 1996  
 A;Title: Extensive chromosome aberrations in Spiroplasma citri strain BR3.  
 A;Reference number: 220500; MUID:97049319; PMID:8894049  
 A;Accession: T28662  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-464 <YEF>  
 A;Cross-references: UNIPROT:Q53886; UNIPARC:UPI000000B9BCD; EMBL:U44405; NID:gl322235; PID:140798  
 C;Genetics:  
 C;Superfamily: Spiroplasma citri hypothetical protein P54

Query Match 45.9%; Score 39; DB 2; Length 464;  
 Best Local Similarity 55.6%; Pred. No. 1.2e+02;  
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 2 WNGESEKPY 10  
 | | | | |  
 Db 364 WDGSGERPY 372

RESULT 41  
 A97200  
 membrane associate histidine kinase with HAMP domain [imported] - Clostridium acetobutylicum  
 C;Species: Clostridium acetobutylicum  
 C;Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
 C;Accession: A97200  
 R;Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
 A;Reference number: A96900; MUID:21359325; PMID:21359325  
 A;Accession: A97200  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-492 <KUR>  
 A;Cross-references: UNIPROT:Q97GD4; UNIPARC:UPI000000CA509; GB:AE001437; PIDN:AAK90388.1; CAC2434  
 A;Experimental source: Clostridium acetobutylicum ATCC824  
 C;Genetics:  
 A;Gene: CAC2434

Query Match 45.9%; Score 39; DB 2; Length 492;  
 Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 QWNGESEKPYDD 12  
 | | | | |  
 Db 56 QYNANENNPYDD 67

RESULT 42  
 I40798  
 cellulase (EC 3.2.1.4) celsa - Clostridium longisporum  
 N;Alternate names: endo-1,4-beta-D-glucanase  
 C;Species: Clostridium longisporum  
 C;Date: 04-Sep-1997 #sequence\_revision 04-Sep-1997 #text\_change 09-Jul-2004  
 C;Accession: I40798  
 R;Mittendorf, V.; Thomson, J.A.  
 J. Gen. Microbiol. 139, 3233-3242, 1993  
 A;Title: Cloning of an endo-(1->4)-beta-glucanase gene, celsa, from the rumen bacterium Clostridium longisporum  
 A;Reference number: I40798; MUID:94172316; PMID:8126442  
 A;Accession: I40798  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-517 <RES>  
 A;Cross-references: UNIPROT:P54937; UNIPARC:UPI0000012BDC9; GB:L02868; NID:gl144754; PIDN:140798  
 C;Genetics:  
 A;Gene: celsa

C;Superfamily: endoglucanase; bacterial cellulose-binding domain homology  
 C;Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation

Query Match 45.9%; Score 39; DB 2; Length 517;  
 Best Local Similarity 60.0%; Pred. No. 1.4e+02;  
 Matches 9; Conservative 0; Mismatches 4; Indels 2; Gaps 1;

Qy 2 WNGESE--KPYDDHL 14  
 | | | | |  
 Db 164 WTQISEAFKDYDDHL 178

RESULT 43  
 T12820  
 hypothetical protein yonF - Bacillus subtilis phage SPBc2  
 C;Species: Bacillus subtilis phage SPBc2  
 C;Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 09-Jul-2004  
 C;Accession: T12820; H69913  
 R;Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mauel, C.; Karamata, D.  
 submitted to the EMBL Data Library, August 1997  
 A;Description: The complete nucleotide sequence of the Bacillus subtilis SPBc2 prophage  
 A;Reference number: Z17583  
 A;Accession: T12820  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: DNA  
 A;Residues: 1-589 <LAZ>  
 A;Cross-references: UNIPROT:O64069; UNIPARC:UPI00000605B8; EMBL:AF020713; NID:g3025478; R;Kunst, F.; Ogatawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Berter, C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho, A.; Ehrlich, S.D.; Emerson, P.T.; Etian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Nature 390, 249-256, 1997  
 A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier, I.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M.F.; Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, A.; Muth, S.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mauel, Y.; M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A.; Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, E.  
 A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
 A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.  
 A;Reference number: A69580; MUID:9804033; PMID:9384377  
 A;Accession: H69913  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-589 <KUN>  
 A;Cross-references: UNIPARC:UPI00000605B8; GB:299115; GB:AL009126; NID:g2634478; PIDN:CA  
 A;Experimental source: strain 168  
 C;Genetics:  
 A;Gene: yonF

Query Match 45.9%; Score 39; DB 2; Length 589;  
 Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 WNGESEKPY 10  
 | | | | |  
 Db 295 WFGSEKAY 303

RESULT 44  
 G72420  
 oligopeptide ABC transporter, periplasmic oligopeptide-binding protein - Thermotoga maritima  
 C;Species: Thermotoga maritima  
 C;Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
 C;Accession: G72420  
 R;Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; C.M.  
 Nature 399, 323-329, 1999  
 A;Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome sequencing  
 A;Reference number: A72200; MUID:99287316; PMID:10360571

A;Accession: G72420  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-625 <ARN>  
A;Cross-references: UNIPROT:Q9WXS6; UNIPARC:UPI000000D3AA6; GB:AE001694; GB:AE000512; NID  
A;Experimental source: strain MSB8  
C;Genetics:  
A;Gene: TM0071  
C;Superfamily: dipeptide transport protein

Query Match 45.9%; Score 39; DB 2; Length 625;  
Best Local Similarity 77.8%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 NGESEKPYD 11  
Db 367 NGEKFPYD 375  
||| |||  
||| |||

RESULT 45  
AH2913  
acetyl-coenzyme A synthetase acs [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C;Species: Agrobacterium tumefaciens  
C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 09-Jul-2004  
C;Accession: AH2913  
R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClellan,  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A;Reference number: AB2577; MUID:21608550; PMID:11743193  
A;Accession: AH2913  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-657 <KUR>  
A;Cross-references: UNIPROT:Q8UBV5; UNIPARC:UPI000000D1FE8; GB:AE008688; PIDN:AAU43726.1;  
A;Experimental source: strain C58 (Dupont)  
C;Genetics:  
A;Gene: acs  
A;Map position: circular chromosome  
C;Superfamily: acetate-CoA ligase; acetate-CoA ligase homology

Query Match 45.9%; Score 39; DB 2; Length 657;  
Best Local Similarity 50.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 2 WNGESEKPYDDH 13  
Db 476 WPGQSRSVYGDH 487  
||| |||  
||| |||

RESULT 46  
C97688  
acs [acetyl-CoA synthetase] gene homolog [imported] - Agrobacterium tumefaciens (strain C58)  
C;Species: Agrobacterium tumefaciens  
C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004  
C;Accession: C97688  
R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
Science 294, 2323-2328, 2001  
A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A;Reference number: A97359; MUID:21608551; PMID:11743194  
A;Accession: C97688  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-657 <KUR>  
A;Cross-references: UNIPROT:Q8UBV5; UNIPARC:UPI000000D1FE8; GB:AE007869; PIDN:AAK88460.1;  
C;Genetics:  
A;Gene: AGR\_C\_4980  
A;Map position: circular chromosome  
C;Superfamily: acetate-CoA ligase; acetate-CoA ligase homology

Query Match 45.9%; Score 39; DB 2; Length 657;  
Best Local Similarity 50.0%; Pred. No. 1.8e+02;  
Matches 6; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 2 WNGESEKPYDDH 13  
Db 476 WPGQSRSVYGDH 487  
||| |||  
||| |||

RESULT 47  
S63993  
acrosomal protein AZ1 - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 20-Jul-1996 #sequence\_revision 08-Nov-1996 #text\_change 09-Jul-2004  
C;Accession: S63993; S63992  
R;Aoto, H.; Tsuchida, J.; Nishina, Y.; Nishimune, Y.; Asano, A.; Tajima, S.  
submitted to the EMBL Data Library, December 1994  
A;Description: Isolation of a novel cDNA that encodes a protein localized to the pre-acrosome  
A;Reference number: S63993  
A;Accession: S63993  
A;Molecule type: mRNA  
A;Residues: 1-1060 <AOT>  
A;Cross-references: UNIPROT:Q62036; UNIPARC:UPI00000285E9; EMBL:D43921; NID:g1827501; PII  
A;Note: this is a revision to the sequence from reference S63992  
R;Aoto, H.; Tsuchida, J.; Nishina, Y.; Nishimune, Y.; Asano, A.; Tajima, S.  
Eur. J. Biochem. 234, 8-15, 1995  
A;Title: Isolation of a novel cDNA that encodes a protein localized to the pre-acrosome  
A;Reference number: S63992; MUID:96096715; PMID:8529672  
A;Accession: S63992  
A;Molecule type: mRNA  
A;Residues: 1-942, 'SGCSP' <AOW>  
A;Cross-references: UNIPARC:UPI000017C61B; EMBL:D43921  
A;Note: this sequence has been revised in reference S63993

Query Match 45.9%; Score 39; DB 2; Length 1060;  
Best Local Similarity 46.2%; Pred. No. 3.1e+02;  
Matches 6; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 QWNGESEKPYDDH 13  
Db 405 QWQSPDKPQDIH 417  
||| |||  
||| |||

RESULT 48  
T33606  
hypothetical protein F54C4.3 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T33606  
R;Rohlfing, T.; Antoniou, B.  
submitted to the EMBL Data Library, October 1998  
A;Description: The sequence of C. elegans cosmid F54C4.  
A;Reference number: Z21376  
A;Accession: T33606  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-1145 <ROH>  
A;Cross-references: UNIPROT:Q9TZ92; UNIPARC:UPI000004E826; EMBL:AF099916; PIDN:AAAC68776.1;  
A;Experimental source: strain Bristol N2; clone F54C4  
C;Genetics:  
A;Gene: CESP:F54C4.3  
A;Map position: 3  
A;Introns: 57/3; 98/1; 167/3; 193/1; 264/3; 307/3; 357/2; 371/2; 432/1; 515/3; 564/2; 632/1;  
C;Superfamily: Caenorhabditis elegans hypothetical protein F54C4.3

Query Match 45.9%; Score 39; DB 2; Length 1145;  
Best Local Similarity 57.1%; Pred. No. 3.4e+02;  
Matches 8; Conservative 1; Mismatches 1; Indels 4; Gaps 1;

Qy 1 QWNGESEKPYDDH 14  
Db 1077 QWSGE----YDDEL 1086  
||| |||  
||| |||

## RESULT 49

SAZQK1  
 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (std)  
 C;Species: Plasmodium falciparum  
 C;Date: 30-Sep-1987 #sequence\_revision 30-Sep-1987 #text\_change 31-Dec-2004  
 C;Accession: A25120  
 R;Mackay, M.; Goman, M.; Bone, N.; Hyde, J.E.; Scaife, J.; Certa, U.; Stunnenberg, H.; E  
 EMBO J. 4, 3823-3829, 1985  
 A;Title: Polymorphism of the precursor for the major surface antigens of Plasmodium falc  
 A;Reference number: A91030; MUID:86136024; PMID:3004972  
 A;Accession: A25120  
 A;Molecule type: DNA  
 A;Residues: 1-1631 <MAC>  
 A;Cross-references: UNIPARC:UPI0000174695  
 C;Comment: The merozoite stages of different strains have strain-specific surface antigen  
 C;Comment: P. falciparum has three stages: sporozoite, merozoite, and gametocyte. The me  
 C;Superfamily: G surface protein  
 C;Keywords: glycoprotein; merozoite; surface antigen; tandem repeat; transmembrane prote  
 F;1-19/Domain: signal sequence #status predicted <SIG>  
 F;20-1631/Product: major merozoite surface antigen #status predicted <MAT>  
 F;67-84/Region: 3-residue repeats (S-G-T/P)  
 F;1614-1631/Domain: membrane anchor #status predicted <MBN>  
 F;97,259,755,759,835,911,955,1049,1156,1165,1436,1563/Binding site: carbohydrate (Asn) (

Query Match 45.9%; Score 39; DB 1; Length 1631;

Best Local Similarity 77.8%; Pred. No. 5e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 GESEKPYDD 12

Db 1238 GESEEDYDD 1246

## RESULT 50

S05603  
 major merozoite surface antigen precursor - malaria parasite (Plasmodium falciparum) (std)  
 N;Alternate names: gp195 surface antigen  
 C;Species: Plasmodium falciparum  
 C;Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 31-Dec-2004  
 C;Accession: S05603; S04850  
 R;Myler, P.J.  
 submitted to the EMBL Data Library, April 1989  
 A;Reference number: S05603  
 A;Accession: S05603  
 A;Molecule type: mRNA  
 A;Residues: 1-1639 <MYL>  
 A;Cross-references: UNIPROT:P04933; UNIPARC:UPI0000000672; EMBL:X15063; NID:g9896; PIDN:  
 R;Myler, P.J.  
 Nucleic Acids Res. 17, 5401, 1989  
 A;Title: Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from Plasm  
 A;Reference number: S04850; MUID:89345116; PMID:2668887  
 A;Accession: S04850  
 A;Molecule type: mRNA  
 A;Residues: 1504-1639 <MYL2>  
 A;Cross-references: UNIPARC:UPI0000177F84; EMBL:X15063  
 C;Superfamily: G surface protein  
 C;Keywords: glycoprotein; merozoite; surface antigen  
 F;1-19/Domain: signal sequence #status predicted <SIG>  
 F;20-1639/Product: major merozoite surface antigen #status predicted <MAT>

Query Match 45.9%; Score 39; DB 2; Length 1639;

Best Local Similarity 77.8%; Pred. No. 5e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 GESEKPYDD 12

Db 1247 GESEEDYDD 1255

Search completed: February 10, 2006, 23:35:15

Job time : 89 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:26:22 ; Search time 252 Seconds  
(without alignments)  
39.196 Million cell updates/sec

Title: US-10-797-821-28

Perfect score: 85

Sequence: 1 QWNGSEKPYDDHL 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85	100.0	1590	2 Q55263_9STRE	Q55263 streptococc
2	85	100.0	1590	2 Q59983_9STRE	Q59983 streptococc
3	85	100.0	1592	1 GTF2_STRDO	P27470 streptococc
4	85	100.0	1597	1 GTF1_STRDO	P11001 streptococc
5	67	78.8	1455	1 GTFC_STRMU	P13470 streptococc
6	67	78.8	1476	1 GTFB_STRMU	P08987 streptococc
7	55	64.7	460	2 Q55MV2_CRYNE	Q55MV2 streptococc
8	55	64.7	460	2 Q5KB79_CRYNE	Q5KB79 streptococc
9	51	60.0	1330	2 Q84CN4_LEUME	Q84CN4 leuconostoc
10	51	60.0	1477	2 Q9L466_LEUME	Q9L466 leuconostoc
11	51	60.0	1508	2 Q9E2H5_LEUME	Q9E2H5 leuconostoc
12	51	60.0	1508	2 Q52224_LEUME	Q52224 leuconostoc
13	48	56.5	105	1 METJ_PHOLL	Q7MYC7 photorhabdu
14	48	56.5	105	1 METJ_YERPE	Q8ZJ18 yersinia pe
15	48	56.5	105	2 Q66G78_YERPS	Q66G78 yersinia ps
16	48	56.5	105	2 Q6CZA0_ERWCT	Q6CZA0 erwinia car
17	48	56.5	222	2 Q4TIA0_TETNG	Q4TIA0 tetraodon n
18	48	56.5	663	2 Q4SS91_TETNG	Q4SS91 tetraodon n
19	48	56.5	754	2 Q4T6C1_TETNG	Q4T6C1 tetraodon n
20	48	56.5	1271	1 Q6BG99_PARTE	Q6BG99 paramecium
21	47	55.3	104	1 METJ_SHEON	Q8EA52 shewanella
22	47	55.3	105	2 Q6LVH6_PHOPR	Q6LVH6 photobacter
23	47	55.3	1463	2 Q5SBM6_LACPE	Q5SBM6 lactobacill
24	47	55.3	1819	2 Q4QUG4_LEIMA	Q4QUG4 leishmania
25	47	55.3	2275	2 Q8IBX2_PLAF7	Q8IBX2 plasmodium
26	46	54.1	166	2 Q56X20_ARATH	Q56X20 arabidopsis
27	46	54.1	424	2 Q4WTQ5_ASPTU	Q4WTQ5 aspergillus
28	46	54.1	427	2 Q93YN4_ARATH	Q93YN4 arabidopsis
29	46	54.1	431	2 Q9FMR7_ARATH	Q9FMR7 arabidopsis
30	46	54.1	899	2 Q97296_PLAF7	Q97296 plasmodium
31	45	52.9	105	1 METJ_VIBCH	Q9KNP9 vibrio chol

32	45	52.9	105	1 METJ_VIBVU	Q8DCN7 vibrio vuln
33	45	52.9	105	1 METJ_VIBVY	Q7MH63 vibrio vuln
34	45	52.9	106	1 METJ_VIBPA	Q87L49 vibrio para
35	45	52.9	373	2 Q7KW8_DICDI	Q7KW8 dictyosteli
36	45	52.9	415	1 SARH_SULTO	Q97500 sulfobolus
37	45	52.9	415	2 Q4JAZ7_SULAC	Q4JAZ7 sulfobolus
38	45	52.9	1187	2 Q6NXJ0_MOUSE	Q6NXJ0 mus musculu
39	44	51.8	85	2 Q71964_9GEMI	Q71964 sida golden
40	44	51.8	107	2 Q5QV41_IDILO	Q5QV41 idiomarina
41	44	51.8	170	2 Q82BK4_STRCO	Q82BK4 streptomyce
42	44	51.8	175	2 Q82LU6_STRAW	Q82LU6 streptomyce
43	44	51.8	292	2 Q6BXN6_DEBHA	Q6BXN6 debaryomyce
44	44	51.8	378	2 Q4RLR4_TETNG	Q4RLR4 tetraodon n
45	44	51.8	411	2 Q874K7_9BASI	Q874K7 rhodospori
46	44	51.8	416	2 Q54549_HALSA	Q54549 halobacteri
47	44	51.8	453	2 Q55JV8_CRYNE	Q55JV8 cryptococcu
48	44	51.8	453	2 Q5K9P2_CRYNE	Q5K9P2 cryptococcu
49	44	51.8	523	2 Q7UKJ8_RHOBA	Q7UKJ8 rhodopirell
50	44	51.8	669	2 Q6BWQ7_DEBHA	Q6BWQ7 debaryomyce
51	44	51.8	887	2 Q98TV0_AGEPH	Q98TV0 agelaius ph
52	44	51.8	1840	2 Q5RCM7_BRARE	Q5RCM7 brachydanio
53	44	51.8	2447	2 Q13149_FUGRU	Q13149 fugu rubrip
54	44	51.8	2609	2 Q7SG70_NEUCR	Q7SG70 neutrospora
55	43.5	51.2	491	2 Q968M2_9TRYP	Q968M2 trypanosoma
56	43	50.6	102	2 Q65F36_BACLD	Q65F36 bacillus li
57	43	50.6	139	2 Q62ZH5_BACLD	Q62ZH5 bacillus li
58	43	50.6	160	2 Q92R19_RHIME	Q92R19 rhizobium m
59	43	50.6	202	1 T4S1_MESAU	Q49111 mesocricetu
60	43	50.6	202	2 Q60548_MESAU	Q60548 mesocricetu
61	43	50.6	202	2 Q53WX2_MESAU	Q53WX2 mesocricetu
62	43	50.6	279	2 Q63M71_EURPS	Q63M71 burkholderi
63	43	50.6	279	2 Q62D35_BURMA	Q62D35 burkholderi
64	43	50.6	322	2 Q5V578_ORYSA	Q5V578 oryza sativ
65	43	50.6	334	2 Q04695_PEA	Q04695 pisum sativ
66	43	50.6	340	2 Q9FXJ4_ARATH	Q9FXJ4 arabidopsis
67	43	50.6	343	2 Q9FS02_ASTSI	Q9FS02 astragalus
68	43	50.6	347	2 Q04696_PEA	Q04696 pisum sativ
69	43	50.6	362	2 Q9UI22_CAEEL	Q9UI22 caenorhabdi
70	43	50.6	373	2 Q4T2P4_TETNG	Q4T2P4 tetraodon n
71	43	50.6	422	2 Q5V579_ORYSA	Q5V579 oryza sativ
72	43	50.6	516	2 Q4HZ71_GIBZE	Q4HZ71 gibberella
73	43	50.6	543	2 Q7XSA5_ORYSA	Q7XSA5 oryza sativ
74	43	50.6	599	2 Q70AT7_HORVU	Q70AT7 hordeum vul
75	43	50.6	604	2 Q6L4X6_ORYSA	Q6L4X6 oryza sativ
76	43	50.6	622	2 Q4XLG6_PLACH	Q4XLG6 plasmodium
77	43	50.6	645	2 Q67SQ4_SYMTH	Q67SQ4 symbiobacte
78	43	50.6	732	2 Q5KGW2_CRYNE	Q5KGW2 cryptococcu
79	43	50.6	900	2 Q4MYQ3_THERA	Q4MYQ3 theileria p
80	43	50.6	910	2 Q6ARD6_DESPS	Q6ARD6 desulfotale
81	43	50.6	1279	2 Q6BSF7_DEBHA	Q6BSF7 debaryomyce
82	43	50.6	1290	2 Q48756_LEUME	Q48756 leuconostoc
83	43	50.6	4767	2 Q17301_CAEBR	Q17301 caenorhabdi
84	43	50.6	7743	2 Q618F4_CAEBR	Q618F4 caenorhabdi
85	42.5	50.0	178	2 Q8SYH4_DROME	Q8SYH4 drosophila
86	42.5	50.0	178	2 Q8INP5_DROME	Q8INP5 drosophila
87	42.5	50.0	268	2 Q9VHG3_DROME	Q9VHG3 drosophila
88	42.5	50.0	414	2 Q6AY84_RAT	Q6AY84 rattus norv
89	42.5	50.0	435	2 Q8EMJ3_OCEIH	Q8EMJ3 oceanobacil
90	42.5	50.0	570	2 Q9YIL2_HETGL	Q9YIL2 heterodera
91	42.5	50.0	1489	2 Q8PSK6_METWA	Q8PSK6 methanosarc
92	42	49.4	104	1 METJ_ECO57	Q8A8U8 escherichia
93	42	49.4	104	1 METJ_ECOL6	Q8A8U7 escherichia
94	42	49.4	104	1 METJ_ECOLI	Q8A8U6 escherichia
95	42	49.4	104	1 METJ_SALTY	Q82224 salmonella
96	42	49.4	104	1 METJ_SALTY	Q82224 salmonella
97	42	49.4	104	1 METJ_SHIFL	Q82224 salmonella
98	42	49.4	105	2 Q57HB6_SALCH	Q57HB6 salmonella
99	42	49.4	105	2 Q5PK48_SALPA	Q5PK48 salmonella
100	42	49.4	106	2 Q5E2I3_VIBF1	Q5E2I3 vibrio fisc
101	42	49.4	125	2 Q9HV11_PSEAE	Q9HV11 pseudomonas
102	42	49.4	135	2 Q99DX6_9RETR	Q99DX6 equine infe
103	42	49.4	187	2 Q54U14_DICDI	Q54U14 dictyosteli
104	42	49.4	214	2 Q4IA43_GIBZE	Q4IA43 gibberella

105	42	49.4	225	2	Q87N92_VIBPA	Q87N92_vibrio para	178	40.5	47.6	1612	2	Q5WMP5_ORYSA	Q5WMP5_oryza sativ
106	42	49.4	280	2	Q9E255_CORUL	Q9e255 corynebacte	179	40	47.1	174	2	Q6K798_ORYSA	Q6k798_oryza sativ
107	42	49.4	300	2	Q6U630_KLEPN	Q6u630 klebsiella	180	40	47.1	179	2	Q44993_CAEEL	Q44993_caenorhabdi
108	42	49.4	317	2	Q73D81_BACC1	Q73d81 bacillus ce	181	40	47.1	180	2	Q4K9X9_PSEF5	Q4k9x9_pseudomonas
109	42	49.4	424	2	Q9FD55_STRN6	Q9fd55 streptococc	182	40	47.1	187	2	Q5E0M3_VIBF1	Q5e0m3_vibrio fisc
110	42	49.4	434	2	Q8DNS5_STRR6	Q8dns5 streptococc	183	40	47.1	203	2	Q4SEQ3_TETNG	Q4seq3_tetradodon n
111	42	49.4	424	2	Q97PB4_STRN9	Q97pb4 streptococc	184	40	47.1	206	2	Q8G4A7_BIFLO	Q8g4a7_bifidobacte
112	42	49.4	479	2	Q4Q5T4_LEIMA	Q4q5t4 leishmania	185	40	47.1	213	2	Q6CPB0_KLULA	Q6cpb0_kluyveromyc
113	42	49.4	489	2	Q6EM13_XENLA	Q6em13 xenopus lae	186	40	47.1	242	1	FSRB_ENTFA	Q9rsg4_enterococcu
114	42	49.4	490	2	Q7RE36_FLAYO	Q7re36 plasmodium	187	40	47.1	258	2	Q83Y51_9BACT	Q83y51_uncultured
115	42	49.4	562	2	Q55134_CRYNE	Q55134 cryptococcu	188	40	47.1	276	2	Q51AU3_ENTHI	Q51au3_entamoeba h
116	42	49.4	573	2	Q43826_RAPSA	Q43826 raphanus sa	189	40	47.1	278	2	Q6C2U6_YARLI	Q6c2u6_yarrowia li
117	42	49.4	577	2	Q620T1_CAEER	Q620t1 caenorhabdi	190	40	47.1	279	2	Q511F7_MAGGR	Q511f7_magnaporthe
118	42	49.4	593	2	Q51M20_MAGGR	Q51m20 magnaporthe	191	40	47.1	303	2	Q76EG2_9EUGL	Q76eg2_eutreptiell
119	42	49.4	602	2	Q87849_STRCO	Q87849 streptomyc	192	40	47.1	307	2	Q8XKD9_CLOPE	Q8xkd9_clostridium
120	42	49.4	736	2	Q4QF12_LEIMA	Q4qf12 leishmania	193	40	47.1	310	2	Q5L927_BACFN	Q5l927_bacteroides
121	42	49.4	774	2	Q5K7M1_CRYNE	Q5k7m1 cryptococcu	194	40	47.1	310	2	Q64P87_BACFR	Q64p87_bacteroides
122	42	49.4	802	2	Q6P154_XENLA	Q6p154 xenopus lae	195	40	47.1	325	2	Q5CM59_CRYHO	Q5cm59_cryptospori
123	42	49.4	805	1	HIF1A_XENLA	Q9i8a9 xenopus lae	196	40	47.1	344	2	Q5HXN2_GLUOX	Q5hxn2_glucanobact
124	42	49.4	980	2	Q8ABA4_BACTN	Q8aba4 bacteroides	197	40	47.1	355	2	Q4L4Z7_STAHL	Q4l4z7_staphylococ
125	42	49.4	1429	1	L1N12_CAEEL	Q14585 caenorhabdi	198	40	47.1	383	2	Q60WV6_CAEER	Q60wv6_caenorhabdi
126	42	49.4	1522	2	Q6TXV4_LEUME	Q6txv4 leuconostoc	199	40	47.1	384	2	Q7V3H1_PROMP	Q7v3h1_prochloroco
127	42	49.4	1527	2	Q8KRE1_LEUME	Q8kre1 leuconostoc	200	40	47.1	385	2	Q8T293_DICD1	Q8t293_diclyostell
128	42	49.4	1527	2	Q9ZAR4_LEUME	Q9zar4 leuconostoc	201	40	47.1	395	2	Q57XZ9_9TRYP	Q57x29_trypanosoma
129	42	49.4	1598	2	Q60NF5_CAEER	Q60nf5 caenorhabdi	202	40	47.1	402	2	Q9FKJ2_ARATH	Q9fkj2_arabidopsis
130	42	49.4	1661	2	Q4N954_THEPA	Q4n954 theileria p	203	40	47.1	407	2	Q5HZ38_ARATH	Q5hz38_arabidopsis
131	41.5	48.8	322	2	Q6YV21_ORYSA	Q6yv21_oryza sativ	204	40	47.1	407	2	Q8LPG6_ARATH	Q8lpg6_arabidopsis
132	41.5	48.8	405	2	Q9NCS4_ENTIV	Q9ncs4 entamoeba i	205	40	47.1	410	2	Q8A5Q2_BACTN	Q8a5q2_bacteroides
133	41.5	48.8	447	2	Q03927_9CAUD	Q03927 bacterioph	206	40	47.1	416	2	Q8VNM6_GLUOX	Q8vnm6_glucanobact
134	41.5	48.8	1063	2	Q75H25_ORYSA	Q75h25_oryza sativ	207	40	47.1	422	2	Q6DMN0_TOXGO	Q6dmn0_toxoplasma
135	41	48.2	216	2	Q5IGV2_9CNID	Q5igv2 nematostell	208	40	47.1	436	1	ENGA_STAEP	Q5hp70_staphylococ
136	41	48.2	216	2	Q4PLU3_9CNID	Q4plu3 nematostell	209	40	47.1	436	1	ENGA_STAEP	Q5hp70_staphylococ
137	41	48.2	220	2	Q89HP9_BRAJA	Q89hp9 bradyrhizob	210	40	47.1	439	2	Q6ZLK0_ORYSA	Q6zlk0_oryza sativ
138	41	48.2	225	2	Q86806_STRCO	Q86806 streptomyc	211	40	47.1	441	2	Q57C56_BRUAB	Q57c56_brucella ab
139	41	48.2	226	1	Q82K89_STRAW	Q82k89 streptomyc	212	40	47.1	441	2	Q8FZM7_BRUSU	Q8fzm7_brucella su
140	41	48.2	226	1	OAZ_SCHJP	Q9hfu9 schizosacch	213	40	47.1	457	2	Q8YI89_BRUME	Q8y189_brucella me
141	41	48.2	253	2	Q8NKI2_9AGAR	Q8nki2 tricholoma	214	40	47.1	464	2	Q96ZH7_SULTO	Q96zh7_sulfolobus
142	41	48.2	258	2	Q83V71_9VIBR	Q83v71_vibrio sp.	215	40	47.1	477	2	Q4P8K4_USTWA	Q4p8k4_ustilago ma
143	41	48.2	263	2	Q3CG12_9VIBR	Q93y12_vibrio sp.	216	40	47.1	513	2	Q9VRA1_DROME	Q9vra1_drosophila
144	41	48.2	263	2	Q52EJ8_MAGGR	Q52ej8 magnaporthe	217	40	47.1	532	1	SREP_PENCH	Q92259_penicillium
145	41	48.2	298	2	Q6CGK6_YARLI	Q6cgk6_yarrowia li	218	40	47.1	546	2	Q4Q4U5_LEIMA	Q4q4u5_leishmania
146	41	48.2	308	2	Q872A5_NEUCR	Q872a5 neurospora	219	40	47.1	554	1	G6P11_PSEPK	Q881w9_pseudomonas
147	41	48.2	336	2	Q5DCW2_SCHJA	Q5dcw2 schistosoma	220	40	47.1	554	1	G6P12_PSEPK	Q88dw7_pseudomonas
148	41	48.2	358	2	Q4W1W7_PHYPA	Q4w1w7 physcomitre	221	40	47.1	554	2	Q4ITD1_AZOVI	Q4itd1_azotobacter
149	41	48.2	363	2	Q9LCC8_OCHAN	Q9lcc8 ochrobactru	222	40	47.1	611	2	Q7S7E7_NEUCR	Q7s7e7_neutrospora
150	41	48.2	381	2	Q6Q999_9BACE	Q6q999 bacteroides	223	40	47.1	615	2	Q6C8H5_YARLI	Q6c8h5_yarrowia li
151	41	48.2	389	2	Q9A412_CAUCR	Q9a412 caulobacter	224	40	47.1	701	2	Q4KXK3_XENTR	Q4kxx3_xenopus tro
152	41	48.2	431	2	Q4P829_USTWA	Q4p829 ustilago ma	225	40	47.1	711	2	Q9IBC7_XENLA	Q9ibc7_xenopus lae
153	41	48.2	438	2	Q4W1W9_PHYPA	Q4w1w9 physcomitre	226	40	47.1	723	2	Q5FTK2_GLUOX	Q5ftk2_glucanobact
154	41	48.2	459	2	Q5R045_IDILO	Q5r045 idiomarina	227	40	47.1	739	2	Q8B171_MOUSE	Q8b171_mus musculu
155	41	48.2	488	2	Q7SCJ4_NEUCR	Q7scj4 neurospora	228	40	47.1	785	2	Q7SB64_NEUCR	Q7sb64_neutrospora
156	41	48.2	546	2	Q82164_STRAW	Q82164 streptomyc	229	40	47.1	792	2	Q4RXC8_TETNG	Q4rxc8_tetradodon n
157	41	48.2	551	2	Q6CWG3_KLULA	Q6cwg3 kluyveromyc	230	40	47.1	798	2	Q4IUH7_AZOVI	Q4iuh7_azotobacter
158	41	48.2	561	2	Q6UBES_TRYCR	Q6ube5 trypanosoma	231	40	47.1	812	2	Q4IVB3_AZOVI	Q4ivb3_azotobacter
159	41	48.2	590	2	Q5NDE4_FUGRU	Q5nde4 fugu rubrip	232	40	47.1	863	2	Q9FAC5_ALTSO	Q9fac5_alteromonas
160	41	48.2	625	2	Q4UIV9_THEAN	Q4uiv9 theileria a	233	40	47.1	908	2	Q8Q0E6_METMA	Q8q0e6_methanosarc
161	41	48.2	669	2	Q4S7V1_TETNG	Q4s7v1 tetradodon n	234	40	47.1	919	2	Q4SJ55_TETNG	Q4sj55_tetradodon n
162	41	48.2	813	2	Q9SU24_ARATH	Q9su24 arabidopsis	235	40	47.1	1001	1	RRP44_YEAST	Q08162_saccharomyc
163	41	48.2	828	2	Q60D18_SOLDE	Q60d18 solanum dem	236	40	47.1	1038	2	Q5KRP4_CRYNE	Q5kpr4_cryptococcu
164	41	48.2	901	2	Q5YCK1_FUSOX	Q5yck1 fusarium ox	237	40	47.1	1059	2	Q55ZT5_CRYNE	Q55zt5_cryptococcu
165	41	48.2	904	2	Q86HL7_DICDI	Q86hl7 dictyosteli	238	40	47.1	1110	2	Q6AUT0_ORYSA	Q6aut0_oryza sativ
166	41	48.2	997	2	Q8II28_PLAF7	Q8ii28 plasmodium	239	40	47.1	1122	2	Q84T42_ORYSA	Q84t42_oryza sativ
167	41	48.2	1062	2	Q6ICQ3_KLULA	Q6icq3 kluyveromyc	240	40	47.1	1151	2	Q7S7C6_NEUCR	Q7s7c6_neutrospora
168	41	48.2	1152	2	Q4PBB0_USTWA	Q4pbb0 ustilago ma	241	40	47.1	1946	2	Q68J72_ATIME	Q68j72_apis mellif
169	41	48.2	1202	1	CSK12_HUMAN	Q8wxe0 homo sapien	242	40	47.1	2173	2	Q9U6Z2_DROME	Q9u6z2_drosophila
170	41	48.2	1455	2	Q9AXC3_ANTIHI	Q9axc3 antirrhinum	243	40	47.1	2177	2	Q9VL32_DROME	Q9vl32_drosophila
171	41	48.2	1459	2	Q7MT84_PORGI	Q7mt84 porphyromon	244	40	47.1	2972	2	Q4RT58_TETNG	Q4rtc58_tetradodon n
172	41	48.2	1596	2	Q4SV99_TETNG	Q4sv99 tetradodon n	245	39.5	46.5	124	2	Q4YUC2_PLABE	Q4yuc2_plasmodium
173	41	48.2	1697	2	Q5B167_DROME	Q5b167 drosophila	246	39.5	46.5	414	1	SCRN1_HUMAN	Q12765_homo sapien
174	41	48.2	1953	2	Q5B167_DROME	Q5b167 drosophila	247	39.5	46.5	478	2	Q6PBE9_ACTAD	Q6pbe9_acinetobact
175	40.5	47.6	245	2	Q6CMV5_KLULA	Q6cmv5 kluyveromyc	248	39.5	46.5	573	2	Q4IKJ0_GIBZE	Q4ikj0_gibberella
176	40.5	47.6	249	2	Q9ZQA5_MYCAT	Q9zqa5 mycoplasma	249	39.5	46.5	636	2	Q6CB14_YARLI	Q6cb14_yarrowia li
177	40.5	47.6	626	2	Q5W612_ORYSA	Q5w612_oryza sativ	250	39.5	46.5	690	2	Q8EFZ5_SHEON	Q8efz5_shewanella

251	39.5	46.5	865	2	Q84QH1_AEGTA	Q84qh1 aegilops ta	324	39	45.9	702	2	Q55UJ1_CRYNE	Q55uj1 cryptococcu
252	39.5	46.5	1208	1	DEC11_DROME	P18169 drosophila	325	39	45.9	702	2	Q5KHV6_CRYNE	Q5khv6 cryptococcu
253	39	45.9	129	2	Q4ZCC9_VIRU	Q4zcc9 bacterioph	326	39	45.9	726	2	Q4X1Z5_ASPFU	Q4x1z5 aspergillus
254	39	45.9	139	2	Q57X53_9TRYP	Q57x53 trypanosoma	327	39	45.9	733	2	Q4N124_THEPA	Q4n124 theileria p
255	39	45.9	168	1	Y407_RHIME	Q92si5 rhizobium m	328	39	45.9	734	2	Q7S1B1_NEUCR	Q7s1b1 neuropept
256	39	45.9	182	2	Q86H75_DICDI	Q86h75 dictyosteli	329	39	45.9	762	2	Q61BU9_CAEBR	Q61bu9 caenorhabdi
257	39	45.9	192	2	Q6PZ03_BAROU	Q6fz03 bartonella	330	39	45.9	766	2	Q9U2C7_CAEBL	Q9u2c7 caenorhabdi
258	39	45.9	237	2	Q8JZK6_MOUSE	Q8jzk6 mus musculu	331	39	45.9	773	2	Q418M2_GIBZE	Q418m2 gibberella
259	39	45.9	247	2	Q5IGV1_9CNID	Q5igv1 nematostell	332	39	45.9	783	2	Q59RK9_CANAL	Q59rk9 candida alb
260	39	45.9	256	2	Q5C3J9_SCHJA	Q5c3j9 schistosoma	333	39	45.9	787	2	Q6SIG0_NAJKA	Q6sig0 najja kaouth
261	39	45.9	256	2	Q4PLU2_9CNID	Q4plu2 nematostell	334	39	45.9	788	2	Q8XGB7_RALSO	Q8xgb7 ralsconia s
262	39	45.9	257	2	Q54CW9_DICDI	Q54cw9 dictyosteli	335	39	45.9	816	2	Q5LN16_SILPO	Q5ln16 silicibacte
263	39	45.9	257	2	Q93G11_9VIBR	Q93g11 vibrio sp	336	39	45.9	841	2	Q9SHQ3_ARATH	Q9shq3 arabidopsis
264	39	45.9	267	2	Q8GDD1_HELMO	Q8gdq1 hellobacill	337	39	45.9	880	2	Q6A7L8_PROAC	Q6a7l8 propionibac
265	39	45.9	275	2	Q510R7_XENTR	Q510r7 xenopus tro	338	39	45.9	896	2	Q8AB23_BACTN	Q8ab23 bacteroides
266	39	45.9	280	2	Q62311_CAEBR	Q62311 caenorhabdi	339	39	45.9	936	2	Q9TT68_MYODA	Q9tt68 myotis daub
267	39	45.9	292	2	Q5HL26_STAEQ	Q5hl26 staphylococ	340	39	45.9	1013	2	Q6CWM3_KLULA	Q6cwm3 kluyveromyc
268	39	45.9	294	2	Q7XDY3_ORYSA	Q7xdy3 oryza sativ	341	39	45.9	1060	1	AZ11_MOUSE	A62036 mus musculu
269	39	45.9	295	2	Q813P5_PLAP7	Q813f5 plasmodium	342	39	45.9	1145	2	Q9T292_CAEBL	Q9t292 caenorhabdi
270	39	45.9	301	2	Q97390_CRAGI	Q97390 crassostrea	343	39	45.9	1603	2	Q54R77_DICDI	Q54r77 dictyosteli
271	39	45.9	303	2	Q9BY31_HUMAN	Q9by31 homo sapien	344	39	45.9	1630	1	MSPI_PLAFW	P04932 plasmodium
272	39	45.9	304	2	Q4WE97_ASPFU	Q4we97 aspergillus	345	39	45.9	1639	1	MSPI_PLAFW	P04933 plasmodium
273	39	45.9	304	2	Q92W06_RHIME	Q92w06 rhizobium m	346	39	45.9	1639	2	Q6LBT0_PLAFA	Q6lbt0 plasmodium
274	39	45.9	309	2	Q8CR16_STAPB	Q8cr16 staphylococ	347	39	45.9	1771	2	Q8TV64_METKA	Q8tv64 methanopyru
275	39	45.9	314	2	Q5AUJ8_EMENI	Q5auj8 aspergillus	348	39	45.9	2112	2	Q8WPL0_9UROC	Q8wpl0 oikopleura
276	39	45.9	322	2	Q7UUE0_RHOBA	Q7uue0 rhodopirell	349	39	45.9	2319	2	Q4RT87_TETNG	Q4rt87 tetradodon n
277	39	45.9	326	2	Q60JT0_CAEBR	Q60jt0 caenorhabdi	350	39	45.9	2327	2	Q91BG7_XENLA	Q91bg7 xenopus lae
278	39	45.9	333	2	Q670V8_PODCA	Q670v8 podocoryne	351	39	45.9	2835	2	Q8G9Q2_LEUME	Q8g9q2 leuconostoc
279	39	45.9	338	2	Q4IU99_AZOV1	Q4iu99 azotobacter	352	39	45.9	3671	2	Q55296_CRYNE	Q55296 cryptococcu
280	39	45.9	339	1	RT13_YEAST	P12686 saccharomyc	353	39	45.9	3671	2	Q5KNL1_CRYNE	Q5knl1 cryptococcu
281	39	45.9	347	2	Q8DYP3_STRAS	Q8dyp3 streptococc	354	38.5	45.3	53	2	Q75VR9_9EURO	Q75vr9 paecilomyce
282	39	45.9	347	2	Q8E4A0_STRA3	Q8e4a0 streptococc	355	38.5	45.3	53	2	Q75V80_9HYPO	Q75v80 paecilomyce
283	39	45.9	351	2	Q8E2M2_STRAW	Q8e2w2 streptomyce	356	38.5	45.3	334	2	Q6AGW0_PROAC	Q6agw0 propionibac
284	39	45.9	355	2	Q8DBP0_VIBVU	Q8dbp0 vibrio vuln	357	38.5	45.3	437	2	Q5WIB8_BACSK	Q5wib8 bacillus cl
285	39	45.9	355	2	Q7MI79_VIBVY	Q7mi79 vibrio vuln	358	38.5	45.3	461	2	Q4R118_CAEBL	Q4r118 caenorhabdi
286	39	45.9	360	2	Q9NAT3_PLAFA	Q9nat3 plasmodium	359	38.5	45.3	512	1	HUTH_BAQU	Q6fz99 bartonella
287	39	45.9	360	2	Q5P7P5_AZOSE	Q5p7p5 azoarcus sp	360	38.5	45.3	535	2	Q55763_SYNY3	Q55763 synechocyst
288	39	45.9	381	2	Q72HQ7_THET2	Q72hq7 thermus the	361	38.5	45.3	555	2	Q17540_CAEBL	Q17540 caenorhabdi
289	39	45.9	387	2	Q8A184_BACTN	Q8a184 bacteroides	362	38.5	45.3	603	2	Q6CL17_KLULA	Q6cl17 kluyveromyc
290	39	45.9	400	2	Q03999_PLAFA	Q03999 plasmodium	363	38.5	45.3	616	2	Q9XIJ3_ARATH	Q9xij3 arabidopsis
291	39	45.9	425	2	Q8IU45_TODPA	Q8iu45 todarodes p	364	38.5	45.3	640	2	Q51KD3_MAGGR	Q51kd3 magnaporthe
292	39	45.9	434	2	Q64EX9_MYCMS	Q64ex9 mycoplasma	365	38.5	45.3	2958	2	Q7RNZ3_PLAYO	Q7rnz3 plasmodium
293	39	45.9	445	2	Q5X081_LEGFL	Q5x081 legionella	366	38	44.7	80	2	Q8X2N0_ECOS7	Q8x2n0 escherichia
294	39	45.9	447	2	Q75NB6_PINFU	Q75nb6 pinetada fu	367	38	44.7	92	2	Q99DS6_9RETR	Q99ds6 equine infe
295	39	45.9	463	2	Q4WDE8_ASPFU	Q4wde8 aspergillus	368	38	44.7	119	2	Q835B0_ENTFA	Q835b0 enterococcu
296	39	45.9	464	2	Q53886_SPICI	Q53886 spiroplasma	369	38	44.7	125	2	Q984Y2_RHILO	Q984y2 rhizobium l
297	39	45.9	465	2	Q55VA4_CRYNE	Q55va4 cryptococcu	370	38	44.7	127	2	Q5WBC3_BACSK	Q5wbc3 bacillus cl
298	39	45.9	468	2	Q4RHAS_TETNG	Q4rhas tetradodon n	371	38	44.7	127	2	Q5WIT0_BACSK	Q5wit0 bacillus cl
299	39	45.9	471	2	Q92FB8_RHIET	Q92fb8 rhizobium e	372	38	44.7	127	2	Q5WJ67_BACSK	Q5wj67 bacillus cl
300	39	45.9	486	2	Q54BE4_DICDI	Q54be4 dictyosteli	373	38	44.7	135	2	Q8AGK9_9RETR	Q8agk9 equine infe
301	39	45.9	489	2	Q5ASB3_EMENI	Q5asb3 aspergillus	374	38	44.7	135	2	Q8AGL1_9RETR	Q8agl1 equine infe
302	39	45.9	492	2	Q97GD4_CLOAB	Q97gd4 clostridium	375	38	44.7	135	2	Q8AGL3_9RETR	Q8agl3 equine infe
303	39	45.9	517	1	GUNA_CLOLO	P54937 clostridium	376	38	44.7	135	2	Q8AGL5_9RETR	Q8agl5 equine infe
304	39	45.9	520	2	Q7UW33_RHOBA	Q7uwv3 rhodopirell	377	38	44.7	135	2	Q8AGL7_9RETR	Q8agl7 equine infe
305	39	45.9	536	2	Q4FXL3_9TRYP	Q4fxl3 trypanosoma	378	38	44.7	135	2	Q8AGL9_9RETR	Q8agl9 equine infe
306	39	45.9	539	2	Q25976_PLAFA	Q25976 plasmodium	379	38	44.7	135	2	Q8AGM1_9RETR	Q8agm1 equine infe
307	39	45.9	539	2	Q25Y91_PLAFA	Q25y91 plasmodium	380	38	44.7	135	2	Q8AGM3_9RETR	Q8agm3 equine infe
308	39	45.9	539	2	Q25971_PLAFA	Q25971 plasmodium	381	38	44.7	135	2	Q8AGM5_9RETR	Q8agm5 equine infe
309	39	45.9	539	2	Q25966_PLAFA	Q25966 plasmodium	382	38	44.7	135	2	Q8AGM7_9RETR	Q8agm7 equine infe
310	39	45.9	539	2	Q25972_PLAFA	Q25972 plasmodium	383	38	44.7	135	2	Q8AGN9_9RETR	Q8agn9 equine infe
311	39	45.9	539	2	Q25981_PLAFA	Q25981 plasmodium	384	38	44.7	135	2	Q8AGN3_9RETR	Q8agn3 equine infe
312	39	45.9	539	2	Q25973_PLAFA	Q25973 plasmodium	385	38	44.7	135	2	Q8AGN5_9RETR	Q8agn5 equine infe
313	39	45.9	539	2	Q25984_PLAFA	Q25984 plasmodium	386	38	44.7	135	2	Q8AGN7_9RETR	Q8agn7 equine infe
314	39	45.9	589	2	Q64069_BPSPC	Q64069 bacterioph	387	38	44.7	135	2	Q8AGN9_9RETR	Q8agn9 equine infe
315	39	45.9	589	2	Q31952_BACSU	Q31952 bacillus su	388	38	44.7	135	2	Q8AGP1_9RETR	Q8agp1 equine infe
316	39	45.9	602	2	Q5B6E0_EMENI	Q5b6e0 aspergillus	389	38	44.7	135	2	Q8AGP3_9RETR	Q8agp3 equine infe
317	39	45.9	617	2	Q4K8E1_PSEF5	Q4k8e1 pseudomonas	390	38	44.7	135	2	Q8AGP5_9RETR	Q8agp5 equine infe
318	39	45.9	624	2	Q4IA87_GIBZE	Q4ia87 gibberella	391	38	44.7	135	2	Q8AGP7_9RETR	Q8agp7 equine infe
319	39	45.9	625	2	Q8GIP4_THENE	Q8gip4 thermotoga	392	38	44.7	135	2	Q8AGP9_9RETR	Q8agp9 equine infe
320	39	45.9	625	2	Q9WXS6_THEMA	Q9wxs6 thermotoga	393	38	44.7	135	2	Q8AGQ1_9RETR	Q8agq1 equine infe
321	39	45.9	627	2	Q95Q23_CABEL	Q95q23 caenorhabdi	394	38	44.7	135	2	Q8AGQ3_9RETR	Q8agq3 equine infe
322	39	45.9	637	2	Q96Q08_DROME	Q96q08 drosophila	395	38	44.7	135	2	Q8AGQ6_9RETR	Q8agq6 equine infe
323	39	45.9	651	1	ACSA_AGRT5	Q8ubv5 agrobacteri	396	38	44.7	135	2	Q8AGR0_9RETR	Q8agr0 equine infe



543	38	44.7	135	2	Q99EG2_9RETR	Q99eg2 equine infe	616	38	44.7	365	2	Q8UAK5_AGR75	Q8uak5 agrobacteri
544	38	44.7	135	2	Q99EG4_9RETR	Q99eg4 equine infe	617	38	44.7	366	2	Q7NSD1_CHRVO	Q7nsd1 chromobacte
545	38	44.7	135	2	Q99EG6_9RETR	Q99eg6 equine infe	618	38	44.7	372	2	Q5TS99_ANOGA	Q5ts99 anopheles g
546	38	44.7	135	2	Q99EG8_9RETR	Q99eg8 equine infe	619	38	44.7	374	2	Q58IJ5_HORVU	Q58ij5 hordeum vul
547	38	44.7	135	2	Q99EH0_9RETR	Q99eh0 equine infe	620	38	44.7	380	1	Q60FTI_PORGI	Q60fti porphyromon
548	38	44.7	135	2	Q99EH2_9RETR	Q99eh2 equine infe	621	38	44.7	380	2	Q60FTI_PORGI	Q60fti porphyromon
549	38	44.7	135	2	Q99EH4_9RETR	Q99eh4 equine infe	622	38	44.7	383	2	Q80564_ARATH	Q80564 arabidopsis
550	38	44.7	135	2	Q99EH6_9RETR	Q99eh6 equine infe	623	38	44.7	386	2	Q80564_ARATH	Q80564 arabidopsis
551	38	44.7	135	2	Q99EH8_9RETR	Q99eh8 equine infe	624	38	44.7	392	2	Q8S6L2_ORYSA	Q8s6l2 oryza sativ
552	38	44.7	135	2	Q99EI0_9RETR	Q99ei0 equine infe	625	38	44.7	393	1	Q8A089_BACTN	Q8a089 bacteroides
553	38	44.7	135	2	Q99EI2_9RETR	Q99ei2 equine infe	626	38	44.7	393	2	Q8LHM4_ORYSA	Q8lhm4 oryza sativ
554	38	44.7	135	2	Q99EI4_9RETR	Q99ei4 equine infe	627	38	44.7	408	2	Q6LDM7_ORYSA	Q6ldm7 oryza sativ
555	38	44.7	135	2	Q99EI6_9RETR	Q99ei6 equine infe	628	38	44.7	408	2	Q8H931_ORYSA	Q8h931 oryza sativ
556	38	44.7	135	2	Q99EU0_9RETR	Q99eu0 equine infe	629	38	44.7	413	2	Q59X42_CANAL	Q59x42 candida alb
557	38	44.7	135	2	Q99EU2_9RETR	Q99eu2 equine infe	630	38	44.7	416	2	Q4HY61_GIBZE	Q4hy61 gibberella
558	38	44.7	135	2	Q99EJ4_9RETR	Q99ej4 equine infe	631	38	44.7	428	2	Q7XF76_ORYSA	Q7xf76 oryza sativ
559	38	44.7	135	2	Q99EJ6_9RETR	Q99ej6 equine infe	632	38	44.7	446	2	Q8S6K3_ORYSA	Q8s6k3 oryza sativ
560	38	44.7	135	2	Q99EK0_9RETR	Q99ek0 equine infe	633	38	44.7	446	2	Q8UJ99_STRHE	Q8uj99 striga herm
561	38	44.7	135	2	Q99EK2_9RETR	Q99ek2 equine infe	634	38	44.7	451	2	Q9LAB9_PSEFL	Q9lab9 pseudomonas
562	38	44.7	135	2	Q99EK6_9RETR	Q99ek6 equine infe	635	38	44.7	461	1	NNT2_CANAL	P46592 candida alb
563	38	44.7	135	2	Q99EK8_9RETR	Q99ek8 equine infe	636	38	44.7	461	2	Q5AJC8_CANAL	Q5ajc8 candida alb
564	38	44.7	135	2	Q99EL0_9RETR	Q99el0 equine infe	637	38	44.7	462	2	Q4HZH7_GIBZE	Q4hzh7 gibberella
565	38	44.7	135	2	Q99EL2_9RETR	Q99el2 equine infe	638	38	44.7	462	1	GBRR2_HUMAN	Q28476 homo sapien
566	38	44.7	148	2	Q6F987_AC1AD	Q6f987 acinetobact	639	38	44.7	465	1	Q8T2W8_TRYCR	Q8t2w8 trypanosoma
567	38	44.7	155	2	Q4YM42_PLABE	Q4ym42 plasmodium	640	38	44.7	475	2	Q5TQ71_ANOGA	Q5tq71 anopheles g
568	38	44.7	156	1	PMP20_SCHPO	O14313 schizosacch	641	38	44.7	485	2	Q96EX8_HUMAN	Q96ex8 homo sapien
569	38	44.7	157	2	Q9XSL9_SHEEP	Q9xsl9 ovis aries	642	38	44.7	488	1	U2AF2_CABBR	P90727 caenorhabdi
570	38	44.7	170	2	Q89JT6_BRAJA	Q89jt6 bradyrhizob	643	38	44.7	492	2	Q9F3D1_STRCO	Q9f3d1 streptomyce
571	38	44.7	183	1	Y793_BRAJA	Q89w99 bradyrhizob	644	38	44.7	495	2	Q9X0H4_THEMA	Q9x0h4 thermotoga
572	38	44.7	194	2	Q9RA50_THETH	Q9ra50 thermus the	645	38	44.7	498	2	Q5AU72_EMENT	Q5au72 aspergillus
573	38	44.7	198	2	Q97IW4_CLOAB	Q97iw4 clostridium	646	38	44.7	501	2	Q9SP93_9ASTR	Q9sp93 matricaria
574	38	44.7	200	2	Q5TYH1_ANOGA	Q5tyh1 anopheles g	647	38	44.7	504	2	Q6I7D2_STAAU	Q6i7d2 staphylococ
575	38	44.7	206	2	Q8B3W2_9GAMA	Q8b3w2 porcine lym	648	38	44.7	506	1	Y6I9_METJA	Q58036 methanococc
576	38	44.7	218	2	Q9W7Y6_9RETR	Q9w7y6 equine infe	649	38	44.7	509	2	Q96NA5_HUMAN	Q96na5 homo sapien
577	38	44.7	224	2	Q66742_9RETR	Q66742 equine infe	650	38	44.7	510	2	Q9U641_DICDI	Q9u641 dictyosteli
578	38	44.7	234	2	Q66749_9RETR	Q66749 equine infe	651	38	44.7	516	2	Q983L0_RHILO	Q983l0 rhizobium l
579	38	44.7	237	2	Q4I016_GIBZE	Q4i016 gibberella	652	38	44.7	547	2	Q07475_STAGA	Q07475 staphylococ
580	38	44.7	243	2	Q5TY12_ANOGA	Q5ty12 anopheles g	653	38	44.7	553	2	Q89F88_BRAJA	Q89f88 bradyrhizob
581	38	44.7	246	2	Q66745_9RETR	Q66745 equine infe	654	38	44.7	579	2	Q7M8H3_WOLUS	Q7m8h3 wolinnella s
582	38	44.7	254	2	Q6Z866_HUMAN	Q6z866 homo sapien	655	38	44.7	589	2	Q9C1I4_FUSSP	Q9c1i4 fusarium sp
583	38	44.7	255	2	Q66750_9RETR	Q66750 equine infe	656	38	44.7	599	2	Q8N9F2_HUMAN	Q8n9f2 homo sapien
584	38	44.7	256	2	Q5E423_VIBFI	Q5e423 vibrio fisc	657	38	44.7	599	2	Q6PIL6_HUMAN	Q6pil6 homo sapien
585	38	44.7	258	2	Q7Y4X2_BPR69	Q7y4x2 bacterioph	658	38	44.7	599	2	Q84YH9_TRIPR	Q84yh9 trifolium p
586	38	44.7	260	2	Q93G21_VIBDI	Q93g21 vibrio diaz	659	38	44.7	602	2	Q6C8I8_YARLI	Q6c8i8 yarrowia li
587	38	44.7	262	2	Q6FJX9_CANGA	Q6fjx9 candida gla	660	38	44.7	609	2	Q76677_CAEEL	Q76677 caenorhabdi
588	38	44.7	267	2	Q5TXH7_ANOGA	Q5txh7 anopheles g	661	38	44.7	615	1	DLLB_BRARE	Q57409 brachydanio
589	38	44.7	282	2	Q8LMH0_ORYSA	Q8lmh0 oryza sativ	662	38	44.7	619	2	Q55NK8_CRYNE	Q55nk8 cryptococcu
590	38	44.7	288	2	Q8LF24_ARATH	Q8lf24 arabidopsis	663	38	44.7	619	2	Q5KB26_CRYNE	Q5kb26 cryptococcu
591	38	44.7	288	2	Q9MI17_ARATH	Q9mi17 arabidopsis	664	38	44.7	623	1	ARP8_BRARE	P59679 brachydanio
592	38	44.7	292	2	Q7WF16_BORER	Q7wf16 bordetella	665	38	44.7	640	2	Q44954_CAEEL	Q44954 caenorhabdi
593	38	44.7	296	2	Q6C9G3_YARLI	Q6c9g3 yarrowia li	666	38	44.7	651	2	Q4ZQGH_PSESY	Q4zqgh pseudomonas
594	38	44.7	296	2	Q8A826_BACTN	Q8a826 bacteroides	667	38	44.7	652	2	Q5KV26_GEOKA	Q5kv26 geobacillus
595	38	44.7	303	2	Q5SK90_THET8	Q5sk90 thermus the	668	38	44.7	653	1	ACSA1_PSEPK	Q88eh6 pseudomonas
596	38	44.7	303	2	Q72KV8_THET2	Q72kv8 thermus the	669	38	44.7	653	2	Q6EMJ3_PSEPU	Q6emj3 pseudomonas
597	38	44.7	303	2	Q5QU67_IDILO	Q5qu67 idiomarina	670	38	44.7	658	2	Q4TPS8_9SPHN	Q4tps8 erythrobact
598	38	44.7	304	2	Q93JU0_PSEST	Q93ju0 pseudomonas	671	38	44.7	674	2	Q5FIW8_ORYSA	Q5fiw8 oryza sativ
599	38	44.7	308	2	Q75DR6_ASHGO	Q75dr6 ashbya goss	672	38	44.7	679	2	Q6FX38_CANGA	Q6fx38 candida gla
600	38	44.7	309	2	Q9NDN9_HYDMA	Q9ndn9 hydra magni	673	38	44.7	681	2	Q9LIX1_STRCO	Q9lix1 streptomyce
601	38	44.7	310	2	Q6D3E5_ERWCT	Q6d3e5 erwinia car	674	38	44.7	698	2	Q5JUM4_ORYSA	Q5jum4 oryza sativ
602	38	44.7	312	2	Q5TRY8_ANOGA	Q5try8 anopheles g	675	38	44.7	708	2	Q6CUM9_KLUJA	Q6cum9 kluyveromyc
603	38	44.7	314	2	Q75BD4_ASHGO	Q75bd4 ashbya goss	676	38	44.7	717	2	Q4XSI4_PLACH	Q4xsi4 plasmodium
604	38	44.7	324	2	Q978Y5_THEVO	Q978y5 thermoplas	677	38	44.7	720	2	Q36332_9RETR	Q36332 equine infe
605	38	44.7	326	2	Q50YG9_ENTHI	Q50yg9 entamoeba h	678	38	44.7	730	2	Q4XEX1_PLACH	Q4xex1 plasmodium
606	38	44.7	333	2	Q5CM09_CRYHO	Q5cmg9 cryptospori	679	38	44.7	742	2	Q62DK5_BURMA	Q62dk5 burkholderi
607	38	44.7	334	2	Q6MHV7_BDEBA	Q6mhv7 bdellovibri	680	38	44.7	762	2	Q63MX4_BURPS	Q63mx4 burkholderi
608	38	44.7	338	2	Q9SHQ2_ARATH	Q9shq2 arabidopsis	681	38	44.7	762	2	Q4V2T5_BURMA	Q4v2t5 burkholderi
609	38	44.7	340	2	Q526U7_WAGGR	Q526u7 magnaporth	682	38	44.7	762	2	Q63MX5_BURPS	Q63mx5 burkholderi
610	38	44.7	350	1	GALE2_CYATE	Q526u7 magnaporth	683	38	44.7	779	2	Q93Y21_ARATH	Q93y21 arabidopsis
611	38	44.7	350	2	Q87M55_VIBPA	Q87m59 vibrio para	684	38	44.7	790	2	Q5B6Y9_EMENT	Q5b6y9 aspergillus
612	38	44.7	352	2	Q7YSX8_CRYPV	Q7ysx8 cryptospori	685	38	44.7	793	1	PHK_GLOVI	Q7nlx2 gloeobacter
613	38	44.7	355	2	Q8ER70_OCEIH	Q8er70 oceanobacil	686	38	44.7	802	2	Q4WLR8_ASPFU	Q4wlr8 aspergillus
614	38	44.7	356	2	Q4HXU2_GIBZE	Q4hxu2 gibberella	687	38	44.7	806	2	Q4WH06_ASPFU	Q4wh06 aspergillus
615	38	44.7	357	2	Q7N8Y8_PHOLL	Q7n8y8 photorhabdu	688	38	44.7	809	2	Q4KDP9_PSEF5	Q4kdp9 pseudomonas

689	38	44.7	837	2	Q91UW5_9ZZZZ	Q91UW5 plasmid pip	762	38	44.7	1093	2	Q8LLX4_ORYSA	Q8llx4 oryza sativ
690	38	44.7	842	2	Q992J9_9RETR	Q992j9 equine infe	763	38	44.7	1099	2	O851S1_POHJU	O851s1 photorhabdu
691	38	44.7	844	2	Q9EP44_9RETR	Q9ep44 equine infe	764	38	44.7	1139	2	O52KR3_MOUSE	O52kr3 mus musculu
692	38	44.7	845	2	Q36315_9RETR	Q36315 equine infe	765	38	44.7	1141	2	Q4QBL6_LEIMA	Q4qbl6 leishmania
693	38	44.7	845	2	Q36317_9RETR	Q36317 equine infe	766	38	44.7	1166	2	Q4QBL6_LEIMA	Q4qbl6 leishmania
694	38	44.7	845	2	Q36318_9RETR	Q36318 equine infe	767	38	44.7	1202	2	Q4UDDA_4TEING	Q4udda4 theileria a
695	38	44.7	845	2	Q36319_9RETR	Q36319 equine infe	768	38	44.7	1301	2	Q4T884_4TEING	Q4t884 tetraodon n
696	38	44.7	845	2	Q36320_9RETR	Q36320 equine infe	769	38	44.7	1311	2	Q06307_ALIAC	Q06307 alicyclobac
697	38	44.7	845	2	Q36321_9RETR	Q36321 equine infe	770	38	44.7	1312	1	Q61D5_DROME	Q61d5 drosophila
698	38	44.7	845	2	Q36322_9RETR	Q36322 equine infe	771	38	44.7	1312	1	PIPI_DROME	P25455 drosophila
699	38	44.7	845	2	Q36323_9RETR	Q36323 equine infe	772	38	44.7	1318	2	Q7KTZ7_DROME	O7ktz7 drosophila
700	38	44.7	845	2	Q36324_9RETR	Q36324 equine infe	773	38	44.7	1388	2	Q4TAJ1_4TEING	Q4taj1 tetraodon n
701	38	44.7	845	2	Q36330_9RETR	Q36330 equine infe	774	38	44.7	1411	2	Q4RYQ7_4TEING	Q4ryq7 tetraodon n
702	38	44.7	845	2	Q42056_9RETR	Q42056 equine infe	775	38	44.7	1431	2	Q5QWE7_IDILO	O5qwe7 idionarina
703	38	44.7	846	2	Q9EP39_9RETR	Q9ep39 equine infe	776	38	44.7	1491	1	MAG11_HUMAN	O96q7 h membrane
704	38	44.7	849	2	Q66731_9RETR	Q66731 equine infe	777	38	44.7	1621	2	Q60VU7_CABER	O60vu7 caenorhabdi
705	38	44.7	853	2	O89469_9RETR	O89469 equine infe	778	38	44.7	1643	2	Q6BGE9_PARTE	O6bge9 paramecium
706	38	44.7	858	2	O89473_9RETR	O89473 equine infe	779	38	44.7	1710	2	O8XRM6_CLOPE	O8xrm6 clostridium
707	38	44.7	859	1	ENV_EIAV1	ENV_EIAV1 equine infe	780	38	44.7	1714	2	O55C43_DICDI	O55c43 dictyosteli
708	38	44.7	859	1	ENV_EIAV2	ENV_EIAV2 equine infe	781	38	44.7	2093	2	Q9XZY8_LEIMA	O9xzy8 leishmania
709	38	44.7	859	1	ENV_EIAV3	ENV_EIAV3 equine infe	782	38	44.7	2118	2	Q6C9E3_YARLI	O6c9e3 yarrowia li
710	38	44.7	859	1	ENV_EIAV9	ENV_EIAV9 equine infe	783	38	44.7	2152	2	O5TQO3_ANOGA	O5tqo3 anopheles g
711	38	44.7	859	1	ENV_EIAVC	ENV_EIAVC equine infe	784	38	44.7	2199	2	Q81BB0_PLAF7	Q81bb0 plasmodium
712	38	44.7	859	1	ENV_EIAVW	ENV_EIAVW equine infe	785	38	44.7	2609	2	Q42UG2_PLABE	Q42ug2 plasmodium
713	38	44.7	859	1	ENV_EIAVY	ENV_EIAVY equine infe	786	38	44.7	3082	2	O5CXX5_CRYPV	O5cxx5 cryptospori
714	38	44.7	859	2	O36316_9RETR	O36316 equine infe	787	37.5	44.1	5439	2	Q81586_PLAF7	O81586 plasmodium
715	38	44.7	859	2	O36320_9RETR	O36320 equine infe	788	37.5	44.1	127	2	Q4J823_SULAC	O4j823 sulfolobus
716	38	44.7	859	2	O36321_9RETR	O36321 equine infe	789	37.5	44.1	134	2	Q7NGG7_GLOVI	O7ngg7 gloeobacter
717	38	44.7	859	2	O36322_9RETR	O36322 equine infe	790	37.5	44.1	159	2	Q6KX8_ECOLI	O6kx8 escherichia
718	38	44.7	859	2	O36323_9RETR	O36323 equine infe	791	37.5	44.1	159	2	Q8CVU1_ECOL6	O8cvu1 escherichia
719	38	44.7	859	2	O36324_9RETR	O36324 equine infe	792	37.5	44.1	175	2	Q6VYT4_SYNY3	O6vyr4 synecocyst
720	38	44.7	859	2	O36325_9RETR	O36325 equine infe	793	37.5	44.1	258	2	Q680M5_ARATH	Q680m5 arabidopsis
721	38	44.7	859	2	O36334_9RETR	O36334 equine infe	794	37.5	44.1	266	2	Q612P1_CABER	O612p1 caenorhabdi
722	38	44.7	859	2	O36335_9RETR	O36335 equine infe	795	37.5	44.1	332	2	O8L110_PARZE	O8l110 paracoccus
723	38	44.7	859	2	O36336_9RETR	O36336 equine infe	796	37.5	44.1	351	2	Q67YQ9_ARATH	O67yq9 arabidopsis
724	38	44.7	859	2	O36337_9RETR	O36337 equine infe	797	37.5	44.1	351	2	Q8VZB1_ARATH	O8vzb1 arabidopsis
725	38	44.7	859	2	O36338_9RETR	O36338 equine infe	798	37.5	44.1	352	2	Q4PDD8_USYMA	Q4pdd8 yarrowia li
726	38	44.7	859	2	O36339_9RETR	O36339 equine infe	799	37.5	44.1	409	2	Q6COK3_YARLI	O6cok3 yarrowia li
727	38	44.7	859	2	O36350_9RETR	O36350 equine infe	800	37.5	44.1	411	2	Q7PY76_ANOGA	O7py76 anopheles g
728	38	44.7	859	2	O36351_9RETR	O36351 equine infe	801	37.5	44.1	435	2	O8KFI6_CHLTE	O8kfi6 chlorobium
729	38	44.7	859	2	O36353_9RETR	O36353 equine infe	802	37.5	44.1	462	1	CH11_CANAL	P46876 candida alb
730	38	44.7	859	2	O36354_9RETR	O36354 equine infe	803	37.5	44.1	462	2	Q5AAH2_CANAL	O5aaH2 candida alb
731	38	44.7	859	2	O36355_9RETR	O36355 equine infe	804	37.5	44.1	465	1	DMAB_BALOB	O6x1e1 balansia ob
732	38	44.7	859	2	O36356_9RETR	O36356 equine infe	805	37.5	44.1	466	2	Q4XTY3_PLACH	Q4xyt3 plasmodium
733	38	44.7	859	2	O42060_9RETR	O42060 equine infe	806	37.5	44.1	466	2	Q4P865_USYMA	O4p865 ustilago ma
734	38	44.7	859	2	O92814_9RETR	O92814 equine infe	807	37.5	44.1	553	2	O6EUS8_ORYSA	O6eus8 oryza sativ
735	38	44.7	859	2	Q77MM8_9RETR	Q77MM8 equine infe	808	37.5	44.1	555	2	Q4NSQ9_9DELT	O4nsg9 anaeromyxob
736	38	44.7	859	2	Q99268_9RETR	Q99268 equine infe	809	37.5	44.1	556	2	Q4QJ8_LEIMA	Q4qdj8 leishmania
737	38	44.7	860	1	ENV_EIAV5	ENV_EIAV5 equine infe	810	37.5	44.1	653	2	Q8UOC9_PYRFU	O8uoc9 pyrococcus
738	38	44.7	860	2	O36333_9RETR	O36333 equine infe	811	37.5	44.1	654	2	O5P5W7_AZOSE	O5p5w7 azoarcus sp
739	38	44.7	863	2	Q992J3_9RETR	Q992j3 equine infe	812	37.5	44.1	686	2	O5N5G7_SYNP6	O5n5g7 synecococc
740	38	44.7	880	2	Q7MDT0_VIBVY	Q7mdt0 vibrio vuln	813	37.5	44.1	694	2	O87H65_VIRPA	O87h65 vibrio para
741	38	44.7	881	2	Q8D6X3_VIBVU	Q8d6x3 vibrio vuln	814	37.5	44.1	785	2	Q9VZQ9_DROME	O9vzq9 drosophila
742	38	44.7	886	2	Q414D2_GIBZE	Q414d2 gibberella	815	37.5	44.1	882	2	Q6IRK5_DROME	O6irk5 drosophila
743	38	44.7	888	2	O81225_ARATH	O81225 arabidopsis	816	37.5	44.1	927	2	Q6TKG4_9ACTO	O6tkg4 streptomyce
744	38	44.7	888	2	O80462_ARATH	O80462 arabidopsis	817	37.5	44.1	978	2	Q4HWP8_GIBZE	O4hwp8 gibberella
745	38	44.7	943	2	O52E25_MAGGR	O52e25 magnaporthe	818	37.5	44.1	1087	2	Q4XTR6_PLACH	O4xtr6 plasmodium
746	38	44.7	943	2	Q7X805_ORYSA	Q7x805 oryza sativ	819	37.5	44.1	1467	2	O56CX8_9STRE	O56cx8 streptococc
747	38	44.7	950	2	Q59WD8_CANAL	Q59wd8 candida alb	820	37.5	44.1	1506	2	O8YL10_ANASP	O8yl10 anabaena sp
748	38	44.7	958	2	Q61D20_CABER	Q61d20 caenorhabdi	821	37.5	44.1	3083	2	Q6LEZ3_PLAF7	O6lez3 plasmodium
749	38	44.7	963	2	Q7XH15_ORYSA	Q7xh15 oryza sativ	822	37	43.5	70	1	RL37A_SULSO	O97qz3 sulfolobus
750	38	44.7	963	2	Q9AYH4_ORYSA	Q9ayh4 oryza sativ	823	37	43.5	70	2	Q9Q0D1_HHV2	O9q0d1 human herpe
751	38	44.7	1010	2	Q9XZG6_4TEMA	Q9xzg6 thermotoga	824	37	43.5	73	2	O5E180_VIBF1	O5e180 vibrio fisc
752	38	44.7	1011	2	Q4HXB7_GIBZE	Q4hxb7 gibberella	825	37	43.5	74	2	O614Q6_ORYSA	O614q6 oryza sativ
753	38	44.7	1037	2	Q9Y080_LUCCU	Q9y080 lucilia cup	826	37	43.5	80	2	Q4XJY9_PLACH	O4xjy9 plasmodium
754	38	44.7	1048	2	Q9UA11_MUSDO	Q9ua11 musca domes	827	37	43.5	85	2	Q7T5U7_9GEMI	O7t5u7 tomato chin
755	38	44.7	1070	2	Q4SBE6_4TEING	Q4sbe6 tetraodon n	828	37	43.5	105	2	O02290_CABEL	O02290 caenorhabdi
756	38	44.7	1080	2	Q66PW7_POHJU	Q66pw7 photorhabdu	829	37	43.5	109	2	Q4THZ9_4TEING	Q4thz9 tetraodon n
757	38	44.7	1085	2	Q7XFW4_ORYSA	Q7xfw4 oryza sativ	830	37	43.5	120	2	Q64WI9_BACFR	O64wi9 bacteroides
758	38	44.7	1085	2	Q8S647_ORYSA	Q8s647 oryza sativ	831	37	43.5	128	1	US384_GAHVG	O05102 gallid herp
759	38	44.7	1089	2	Q4N667_4TEPA	Q4n667 theileria p	832	37	43.5	131	2	Q8PD22_XANCP	O8pd22 xanthomonas
760	38	44.7	1090	2	Q7NA40_POHLL	Q7na40 photorhabdu	833	37	43.5	135	2	Q99E88_9RETR	O99e88 equine infe
761	38	44.7	1091	2	Q54CD8_DICDI	Q54cd8 dictyosteli	834	37	43.5	135	2	Q99EK4_9RETR	O99ek4 equine infe

835	37	43.5	137	2	Q8ER15_OCEIH	Q8er15	oceanobacil	908	37	43.5	317	2	Q4QFJ6_LEIMA	Q4qfj6	leishmania
836	37	43.5	138	2	Q4P879_USTWA	Q4p879	ustlago ma	909	37	43.5	321	2	Q9HKX9_THEAC	Q9hkx9	thermoplasm
837	37	43.5	139	2	Q4QIL6_LEIMA	Q4qil6	leishmania	910	37	43.5	322	2	Q57ZB4_9TRYP	Q57zb4	trypanosoma
838	37	43.5	140	2	Q4QJK1_LEIMA	Q4qjk1	leishmania	911	37	43.5	322	2	Q5GY07_XANOR	Q5gy07	xanthomonas
839	37	43.5	141	2	Q61ZL5_PSESM	Q61zl5	pseudomonas	912	37	43.5	322	2	Q6G2C3_BAHE	Q6g2c3	bartonella
840	37	43.5	142	2	Q61ZL6_PSESM	Q61zl6	pseudomonas	913	37	43.5	326	2	Q92KD3_RHIME	Q92kd3	rhizobium m
841	37	43.5	143	2	Q61ZL8_PSESM	Q61zl8	pseudomonas	914	37	43.5	326	2	Q8G3M7_BIFLO	Q8g3m7	bifidobacte
842	37	43.5	149	2	Q61ZL9_PSESZ	Q61zl9	pseudomonas	915	37	43.5	332	2	Q8GGH8_9ENFR	Q8gggh8	enterobacte
843	37	43.5	149	2	Q61ZM1_PSESY	Q61zm1	pseudomonas	916	37	43.5	335	2	Q51TG7_MAGGR	Q51tg7	magnaporthe
844	37	43.5	149	2	Q61ZM4_PSESY	Q61zm4	pseudomonas	917	37	43.5	336	2	Q4J6Z5_SULAC	Q4j6z5	sulfolobus
845	37	43.5	149	2	Q61ZM5_PSESY	Q61zm5	pseudomonas	918	37	43.5	336	2	Q6IR76_XENLA	Q6ir76	xenopus lae
846	37	43.5	149	2	Q61ZM6_PSESY	Q61zm6	pseudomonas	919	37	43.5	338	2	Q8XOM4_NEUCR	Q8xom4	neurospora
847	37	43.5	149	2	Q61ZM7_PSESS	Q61zm7	pseudomonas	920	37	43.5	338	2	Q89Z17_BACTIN	Q89z17	bacteroides
848	37	43.5	149	2	Q61ZM8_PSESY	Q61zm8	pseudomonas	921	37	43.5	342	2	Q4QVH6_9HIVI	Q4qvhl6	human immun
849	37	43.5	149	2	Q61ZM9_PSESK	Q61zm9	pseudomonas	922	37	43.5	342	2	Q4QVJ1_9HIVI	Q4qvjl1	human immun
850	37	43.5	149	2	Q61ZN0_PSESK	Q61zn0	pseudomonas	923	37	43.5	342	2	Q4QVJ7_9HIVI	Q4qvjl7	human immun
851	37	43.5	149	2	Q61ZN5_PSESJ	Q61zn5	pseudomonas	924	37	43.5	342	2	Q4QVP2_9HIVI	Q4qvlp2	human immun
852	37	43.5	149	2	Q61ZN9_PSESH	Q61zn9	pseudomonas	925	37	43.5	342	2	Q4QVP4_9HIVI	Q4qvlp4	human immun
853	37	43.5	149	2	Q61ZP1_PSESD	Q61zp1	pseudomonas	926	37	43.5	342	2	Q4QVP8_9HIVI	Q4qvlp8	human immun
854	37	43.5	149	2	Q61ZP3_PSESD	Q61zp3	pseudomonas	927	37	43.5	342	2	Q4QVR4_9HIVI	Q4qvpr4	human immun
855	37	43.5	149	2	Q61ZP4_PSESD	Q61zp4	pseudomonas	928	37	43.5	342	2	Q4QVR8_9HIVI	Q4qvpr8	human immun
856	37	43.5	149	2	Q61ZP5_PSESD	Q61zp5	pseudomonas	929	37	43.5	346	2	Q5S6W1_DICDI	Q5s6w1	dictyosteli
857	37	43.5	149	2	Q61ZP6_PSESD	Q61zp6	pseudomonas	930	37	43.5	348	2	Q4I7M2_GIBZE	Q4i7m2	gibberella
858	37	43.5	149	2	Q61ZP7_PSEYM	Q61zp7	pseudomonas	931	37	43.5	348	2	Q66B76_YERPS	Q66b76	yersinia ps
859	37	43.5	149	2	Q61ZP9_PSEYM	Q61zp9	pseudomonas	932	37	43.5	349	2	Q4JN55_9MCCC	Q4jnj5	arthrobacte
860	37	43.5	149	2	Q61ZQ5_PSEYM	Q61zq5	pseudomonas	933	37	43.5	351	2	Q9SVD4_ARATH	Q9svd4	arabidopsis
861	37	43.5	149	2	Q61ZQ9_PSESL	Q61zq9	pseudomonas	934	37	43.5	353	2	Q7QA38_ANGGA	Q7vqa38	anopheles g
862	37	43.5	149	2	Q61ZR0_PSESG	Q61zr0	pseudomonas	935	37	43.5	363	2	Q5IUY3_MAGGR	Q5iuy3	magnaporthe
863	37	43.5	149	2	Q61ZR9_PSESG	Q61zr9	pseudomonas	936	37	43.5	363	2	Q8MMP1_9DRIPT	Q8mmp1	anastrepha
864	37	43.5	149	2	Q61ZS0_PSESD	Q61zs0	pseudomonas	937	37	43.5	366	1	DIAC_MOUSE	Q8r242	mus musculu
865	37	43.5	149	2	Q61ZS1_PSESD	Q61zs1	pseudomonas	938	37	43.5	367	1	DIAC_RAT	Q01460	rattus norv
866	37	43.5	149	2	Q61ZS2_PSEAP	Q61zs2	pseudomonas	939	37	43.5	372	2	Q5SQG3_CRYNE	Q5sqg3	cryptococcu
867	37	43.5	149	2	Q61ZS3_PSEAP	Q61zs3	pseudomonas	940	37	43.5	372	2	Q5KFQ4_CRYNE	Q5kfq4	cryptococcu
868	37	43.5	149	2	Q61ZS4_PSESF	Q61zs4	pseudomonas	941	37	43.5	377	2	Q8XSC6_RALSO	Q8xsc6	raistonia s
869	37	43.5	160	2	Q7RUK0_NEUCR	Q7ruk0	neurospora	942	37	43.5	377	2	Q8YWM2_ANASP	Q8ywm2	anabaena sp
870	37	43.5	161	2	Q72DY8_DESVH	Q72dy8	desulfovibr	943	37	43.5	386	2	Q9FCG7_STRCO	Q9fcg7	streptomyce
871	37	43.5	165	2	Q98OM2_SULSO	Q98om2	sulfolobus	944	37	43.5	393	2	Q8KY29_STRCU	Q8ky29	streptomyce
872	37	43.5	165	2	Q4TKB9_9SPHN	Q4tkb9	erythrobact	945	37	43.5	397	2	Q4H917_9DEIO	Q4h917	deinococcu
873	37	43.5	166	1	IFNG_BOSIN	Q86y6	bos indicus	946	37	43.5	400	2	Q66FU0_YERPS	Q66fu0	yersinia ps
874	37	43.5	170	2	Q8DSV5_STRMU	Q8dsv5	streptococc	947	37	43.5	409	2	Q58LL7_9CAUD	Q58ll7	cyanophag
875	37	43.5	174	2	Q8EFY1_SHEON	Q8eff1	shewanella	948	37	43.5	417	2	Q6XPU7_9BRAS	Q6xpu7	leavenworth
876	37	43.5	174	2	Q6ANZ9_DESPS	Q6anz9	desulfotale	949	37	43.5	419	2	Q7MTK7_PORGI	Q7mtk7	porphyromon
877	37	43.5	185	1	SC6_SCRGO	Q74300	schizophyll	950	37	43.5	427	2	Q4T730_TETNG	Q4t730	tetradon n
878	37	43.5	192	2	Q5ST64_HUMAN	Q5st64	homo sapien	951	37	43.5	428	2	Q6FLM1_CANGA	Q6flm1	candida gla
879	37	43.5	193	1	YPMQ_BACSU	P54178	bacillus su	952	37	43.5	428	2	Q502F0_BRARE	Q502f0	brachydanio
880	37	43.5	193	2	Q9S445_PSEAE	Q9s445	pseudomonas	953	37	43.5	431	2	Q987B7_RHILO	Q987b7	rhizobium l
881	37	43.5	193	2	Q6DT23_ENTCL	Q6dt23	enterobacte	954	37	43.5	433	2	Q4HSN2_CAMPUP	Q4hsn2	campylobact
882	37	43.5	194	2	Q4TEF5_TETNG	Q4tef5	tetradon n	955	37	43.5	434	2	Q4KKM6_PSERF	Q4kkm6	pseudomonas
883	37	43.5	195	2	Q738W5_BACCI	Q738w5	bacillus ce	956	37	43.5	437	2	Q5BFA0_EMENI	Q5bfa0	aspergillus
884	37	43.5	209	1	YJDF_ECOLI	P39270	escherichia	957	37	43.5	437	2	Q8LDT6_ARATH	Q8ldt6	arabidopsis
885	37	43.5	209	2	Q83IR5_SHIFL	Q83ir5	shigella fl	958	37	43.5	437	2	Q944K2_ARATH	Q944k2	arabidopsis
886	37	43.5	209	2	Q8XDT2_EC057	Q8xd2	escherichia	959	37	43.5	437	2	Q94CD6_ARATH	Q94cd6	arabidopsis
887	37	43.5	220	2	Q4I2X9_GIBZE	Q4i2x9	gibberella	960	37	43.5	437	2	Q97ID8_CLOAB	Q97id8	clostridium
888	37	43.5	222	2	Q98A52_RHILO	Q98a52	rhizobium l	961	37	43.5	439	2	Q5B3L4_EMENI	Q5b3l4	aspergillus
889	37	43.5	234	1	TRAR_AGRVI	P33909	agrobacteri	962	37	43.5	442	2	Q91VR2_ARATH	Q91vr2	arabidopsis
890	37	43.5	234	2	Q9R486_9RHIZ	Q9r486	agrobacteri	963	37	43.5	445	2	Q51V82_MAGGR	Q51v82	magnaporthe
891	37	43.5	234	2	Q9WMA5_9RHIZ	Q9wma5	agrobacteri	964	37	43.5	445	2	Q9TZ81_CAEEL	Q9tz81	caenorhabdi
892	37	43.5	234	2	Q9WMB2_9RHIZ	Q9wmb2	agrobacteri	965	37	43.5	446	2	Q6CHQ0_YARLI	Q6chq0	yarrowia li
893	37	43.5	235	2	Q4WM48_ASPFU	Q4wm48	aspergillus	966	37	43.5	448	2	Q7X810_ORYSA	Q7x810	oryza sativ
894	37	43.5	238	2	Q4IG90_GIBZE	Q4ig90	gibberella	967	37	43.5	455	1	PEL6_ARATH	Q64510	arabidopsis
895	37	43.5	241	2	Q4TPR8_9SPHN	Q4tpr8	erythrobact	968	37	43.5	455	2	Q5UIE6_ARATH	Q5ui6	arabidopsis
896	37	43.5	242	2	Q8GT53_HORVU	Q8gt53	hordeum vul	969	37	43.5	457	2	Q9HBR1_HUMAN	Q9hbr1	homo sapien
897	37	43.5	243	2	Q67SQ7_SYMTB	Q67sq7	symbiobacte	970	37	43.5	460	2	Q5AWT5_EMENI	Q5awt5	aspergillus
898	37	43.5	250	2	Q5N3D7_SYNP6	Q5n3d7	synchococc	971	37	43.5	460	2	Q5CFS7_CRYHO	Q5cfs7	cryptospori
899	37	43.5	264	2	Q5C399_SCHJA	Q5c399	schistosoma	972	37	43.5	462	2	Q5CSB4_CRYPV	Q5csb4	cryptospori
900	37	43.5	271	2	Q97FP0_CLOAB	Q97fp0	clostridium	973	37	43.5	467	2	Q8TH09_PYRFU	Q8th09	pyrococcus
901	37	43.5	290	2	Q6RVY1_BIFAN	Q6rvy1	bifidobacte	974	37	43.5	467	2	Q7Q265_ANGGA	Q7q265	anopheles g
902	37	43.5	297	2	Q9RD48_STRCO	Q9rd48	streptomyce	975	37	43.5	469	1	SAHH_PSEAE	Q91685	pseudomonas
903	37	43.5	298	2	Q6N1J3_CORDI	Q6n1j3	corynebacte	976	37	43.5	469	1	SAHH_PSESM	Q87v73	pseudomonas
904	37	43.5	308	2	Q4F8N0_LACPL	Q4f8n0	lactobacill	977	37	43.5	469	2	Q4Z292_PSESY	Q4z292	pseudomonas
905	37	43.5	308	2	Q7ZVB9_BRARE	Q7zvb9	brachydanio	978	37	43.5	469	2	Q4K4H7_PSERF	Q4k4h7	pseudomonas
906	37	43.5	309	2	Q9D4Q1_MOUSE	Q9d4q1	mus musculu	979	37	43.5	471	2	Q58704_PYRHO	Q58704	pyrococcus
907	37	43.5	310	2	Q75A04_ASHGO	Q75a04	ashbya goss	980	37	43.5	472	2	Q5ZJMS_CHICK	Q5zjms	gallus gall



981 37 43.5 474 2 Q4FT37\_9GAMM  
 982 37 43.5 475 2 Q7T3Q3\_BRARE  
 983 37 43.5 478 2 Q84SQ3\_ORYSA  
 984 37 43.5 479 2 Q4WX30\_ASPPU  
 985 37 43.5 479 2 Q8T446\_DROME  
 986 37 43.5 479 2 Q9VJN5\_DROME  
 987 37 43.5 481 2 Q6NVR0\_BRARE  
 988 37 43.5 485 2 Q53737\_STRCH  
 989 37 43.5 489 2 Q4HVC9\_9DEIO  
 990 37 43.5 490 2 Q4IUAB\_GIBZE  
 991 37 43.5 493 2 Q7D9U8\_MYCTU  
 992 37 43.5 494 2 Q9PIC5\_HUMAN  
 993 37 43.5 498 2 Q73T14\_MYCPA  
 994 37 43.5 500 2 P96264\_MYCTU  
 995 37 43.5 504 2 Q7U215\_MYCBO  
 996 37 43.5 504 2 Q5AX41\_EMENI  
 997 37 43.5 509 2 Q626D4\_CAEBR  
 998 37 43.5 509 2 Q995C5\_9VIRU  
 999 37 43.5 510 2 Q8B119\_9VIRU  
 1000 37 43.5 514 2 Q4RV71\_TETNG

## ALIGNMENTS

RESULT 1  
 Q5263\_9STRE PRELIMINARY; PRT; 1590 AA.  
 AC Q5263;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE GTF-I.  
 GN Name:Glucosyltransferase;  
 OS Streptococcus sobrinus;  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus;  
 OX NCBI\_TaxID=1310;  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=ATCC 33478;  
 RC Sato S.;  
 RA "DNA and amino-acid sequences of water-insoluble-glucan synthetase  
 RT produced from Streptococcus sobrinus ATCC 33478.";  
 RL Kagoshima Daigaku Shigakubu Kiyo 16:23-29 (1996).  
 RN [2]  
 RP PROTEIN SEQUENCE.  
 RX MEDLINE=91224988; PubMed=1827439;  
 RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;  
 RT "Isolation and sequence of an active-site peptide containing a  
 RT catalytic aspartic acid from two Streptococcus sobrinus alpha-  
 RT glucosyltransferases.";  
 RL J. Biol. Chem. 266:8916-8922 (1991).  
 DR EMBL; D63570; BAA09792.1; -; Genomic\_DNA.  
 DR PIR; A39841; A39841.  
 DR HSSP; P06653; LGVM.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW binding\_1; 3.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6E4FD43 CRC64;  
 Query Match 100.0%; Score 85; DB 2; Length 1590;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDHL 14  
 |||||||||||||  
 Db 337 QWNGESEKPYDDHL 350

RESULT 2

Q59983\_9STRE PRELIMINARY; PRT; 1590 AA.  
 AC Q59983;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Glucosyltransferase-I precursor (EC 2.4.1.5).  
 GN Names:gtfI;  
 OS Streptococcus sobrinus;  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus;  
 OX NCBI\_TaxID=1310;  
 RN NUCLEOTIDE SEQUENCE.  
 RP STRAIN=OMZ176;  
 RC MEDLINE=94146405; PubMed=8312602;  
 RX Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;  
 RA "DNA sequence of the glucosyltransferase gene of serotype d  
 RT Streptococcus sobrinus.";  
 RL DNA Seq. 4:19-27 (1993).  
 RN [2]  
 RP PROTEIN SEQUENCE.  
 RX MEDLINE=91224988; PubMed=1827439;  
 RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;  
 RT "Isolation and sequence of an active-site peptide containing a  
 RT catalytic aspartic acid from two Streptococcus sobrinus alpha-  
 RT glucosyltransferases.";  
 RL J. Biol. Chem. 266:8916-8922 (1991).  
 DR EMBL; D13858; BAA02976.1; -; Genomic\_DNA.  
 DR PIR; A39841; A39841.  
 DR HSSP; P06653; IHX.  
 DR GO; GO:0047849; F:dextranase activity; IEA.  
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW binding\_1; 2.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Glucosyltransferase; Signal; Transferase.  
 FT SIGNAL 1 38 Potential.  
 FT CHAIN 39 1590 Glucosyltransferase-I.  
 SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;  
 Query Match 100.0%; Score 85; DB 2; Length 1590;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDHL 14  
 |||||||||||||  
 Db 337 QWNGESEKPYDDHL 350

RESULT 3  
 GTF2\_STRDO STANDARD; PRT; 1592 AA.  
 AC P27470;  
 DT 01-AUG-1992 (Rel. 23, Created)  
 DT 01-AUG-1992 (Rel. 23, Last sequence update)  
 DT 13-SEP-2005 (Rel. 48, Last annotation update)  
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)  
 DE (sucrose 6-glucosyltransferase).  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus;  
 OX NCBI\_TaxID=1317;  
 RN NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
 RP STRAIN=6715 / Serotype G;  
 RX MEDLINE=91123227; PubMed=1704006;  
 RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,  
 RA Kagawa H.;  
 RT "Peptide sequences for sucrose splitting and glucan binding within  
 RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan



RT synthetase".  
 RL J. Bacteriol. 173:989-996(1991).  
 CC -!- FUNCTION: Production of extracellular glucans, that are thought to  
 CC play a key role in the development of the dental plaque because of  
 CC their ability to adhere to smooth surfaces and mediate the  
 CC aggregation of bacterial cells and food debris.  
 CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-  
 CC fructose + (1,6-alpha-D-glucosyl)(n+1).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha  
 CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes  
 CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both  
 CC forms of glucans.  
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.  
 CC -!- SIMILARITY: Contains 16 cell wall binding repeats.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use as long as its content is in no way modified and this statement is not  
 CC removed.  
 CC -----  
 CC EMBL; D90213; BAA14241.1; -; Genomic DNA.  
 DR HSP; P06653; IGV.  
 DR InterPro; IPR002479; Cell wall bd put.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 3.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.  
 FT SIGNAL 1 38 Potential.  
 FT CHAIN 39 1592 Glucosyltransferase-I.  
 FT REPEAT 1093 1142 1.  
 FT REPEAT 1158 1207 2.  
 FT REPEAT 1222 1272 3.  
 FT REPEAT 1287 1337 4.  
 FT REPEAT 1402 1451 5.  
 FT REPEAT 1514 1563 6.  
 FT REPEAT 1577 1592 7 (incomplete).  
 FT REGION 39 1044 Catalytic (approximate).  
 FT REGION 1093 1592 7 X tandem repeats.  
 FT REGION 1093 1592 Glucan-binding (approximate).  
 FT REGION 1592 1592 Glucan-binding (approximate).  
 SQ SEQUENCE 1592 AA; 176168 MW; BCOA66D079351ECF CRC64;

Query Match 100.0%; Score 85; DB 1; Length 1592;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDHL 14  
 |||||  
 Db 337 QWNGESEKPYDDHL 350

RESULT 4  
 GTFL\_STRDO  
 ID GTFL\_STRDO STANDARD; PRT; 1597 AA.  
 AC P11001;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 13-SEP-2005 (Rel. 48, Last annotation update)  
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)  
 DE (Sucrose 6-glucosyltransferase).  
 GN Name=gtfI;  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1317;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
 RC STRAIN=MFE28;  
 RX MEDLINE=87308014; PubMed=3040686;  
 RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;  
 RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus  
 sobrinus MFE28.";

J. Bacteriol. 169:4271-4278(1987).  
 CC -!- FUNCTION: Production of extracellular glucans, that are thought to  
 CC play a key role in the development of the dental plaque because of  
 CC their ability to adhere to smooth surfaces and mediate the  
 CC aggregation of bacterial cells and food debris.  
 CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-  
 CC fructose + (1,6-alpha-D-glucosyl)(n+1).  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha  
 CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes  
 CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both  
 CC forms of glucans.  
 CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.  
 CC -!- SIMILARITY: Contains 19 cell wall binding repeats.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use as long as its content is in no way modified and this statement is not  
 CC removed.  
 CC -----  
 CC EMBL; M17391; AAC63063.1; -; Genomic DNA.  
 DR InterPro; IPR002479; Cell wall bd put.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 4.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.  
 FT SIGNAL 1 38 Potential.  
 FT CHAIN 39 1597 Glucosyltransferase-I.  
 FT REPEAT 1099 1132 A repeat.  
 FT REPEAT 1163 1213 AC repeat.  
 FT REPEAT 1227 1277 AC repeat.  
 FT REPEAT 1292 1342 AC repeat.  
 FT REPEAT 1352 1399 B repeat.  
 FT REPEAT 1406 1455 AC repeat.  
 FT REPEAT 1465 1512 B repeat.  
 FT REPEAT 1519 1568 AC repeat.  
 FT REPEAT 1582 1597 A repeat (incomplete).  
 FT REGION 39 1050 Catalytic (approximate).  
 FT REGION 1099 1597 1.25 A, 2 B and 5 AC repeats.  
 FT REGION 1099 1597 Glucan-binding (approximate).  
 FT REGION 1597 1597 Glucan-binding (approximate).  
 SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 100.0%; Score 85; DB 1; Length 1597;  
 Best Local Similarity 100.0%; Pred. No. 0.00012;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDHL 14  
 |||||  
 Db 343 QWNGESEKPYDDHL 356

RESULT 5  
 GTFC\_STRMU  
 ID GTFC\_STRMU STANDARD; PRT; 1455 AA.  
 AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;  
 DT 01-NOV-1988 (Rel. 09, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-MAY-2005 (Rel. 47, Last annotation update)  
 DE Glucosyltransferase-SI precursor (EC 2.4.1.5) (GTF-SI)  
 DE (Dextranucrase) (Sucrose 6-glucosyltransferase).  
 GN Name=gtfC; OrderedLocustNames=SMU.1005;  
 OS Streptococcus mutans.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1309;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=GS-5;  
 RX MEDLINE=89137980; PubMed=2976010; DOI=10.1016/0378-1119(88)90382-4;  
 RA Ueda S., Shiroza T., Kuramitsu H.K.;  
 RT "Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.";  
 RL Gene 69:101-109(1988).

[2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,  
 RC MT4467 / Serotype e, and MT8148 / Serotype c;  
 RX MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 RA Kimura S., Hamada S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RT Streptococcus mutans.",  
 RL FEMS Microbiol. Lett. 161:331-336(1998).  
 [3]  
 RN NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=UA159 / ATCC 700610 / Serotype c;  
 RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;  
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,  
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,  
 RA Ferretti J.J.;  
 RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental  
 RT pathogen.",  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
 [4]  
 RN NUCLEOTIDE SEQUENCE OF 1-349.  
 RC STRAIN=GS-5;  
 RX MEDLINE=87308013; PubMed=3040685;  
 RA Shiroza T., Ueda S., Kuramitsu H.K.;  
 RT "Sequence analysis of the gtfB gene from Streptococcus mutans.";  
 RL J. Bacteriol. 169:4263-4270(1987).  
 CC -I- FUNCTION: Production of extracellular glucans, that are thought to  
 CC play a key role in the development of the dental plaque because of  
 CC their ability to adhere to smooth surfaces and mediate the  
 CC aggregation of bacterial cells and food debris.  
 CC -I- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-  
 CC fructose + (1,6-alpha-D-glucosyl)(n+1).  
 CC -I- SUBCELLULAR LOCATION: Secreted.  
 CC -I- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha  
 CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes  
 CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both  
 CC forms of glucans.  
 CC -I- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.  
 CC -I- SIMILARITY: Contains 5 cell wall binding repeats.  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use as long as its content is in no way modified and this statement is not  
 CC removed.  
 CC EMBL; M22054; AAA88592.1; -; Genomic DNA.  
 DR EMBL; D88652; BAA26102.1; -; Genomic DNA.  
 DR EMBL; D88653; BAA26106.1; -; Genomic DNA.  
 DR EMBL; D88654; BAA26110.1; -; Genomic DNA.  
 DR EMBL; D88655; BAA26111.1; -; Genomic DNA.  
 DR EMBL; D88656; BAA26112.1; -; Genomic DNA.  
 DR EMBL; D88657; BAA26113.1; -; Genomic DNA.  
 DR EMBL; D88658; BAA26114.1; -; Genomic DNA.  
 DR EMBL; D88659; BAA26115.1; -; Genomic DNA.  
 DR EMBL; A014940; AAN58706.1; -; Genomic DNA.  
 DR EMBL; M17361; AAA88589.1; -; Genomic DNA.  
 DR PIR; JT0345; JT0345.  
 DR HSSP; P06653; IH8G.  
 DR InterPro; IPR002479; Cell wall bd put.  
 DR InterPro; IPR003318; Glyco hydro 70.  
 DR Pfam; PF01473; CW binding 1; 2.  
 DR Pfam; PF02324; Glyco hydro 70; 1.  
 KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;  
 KW Transferrase.  
 FT SIGNAL 1 34  
 FT CHAIN 35 1455 Glucosyltransferase-SI.  
 FT REPEAT 1126 1159 A repeat.  
 FT REPEAT 1169 1200 A repeat.  
 FT REPEAT 1227 1238 C repeat.  
 FT REPEAT 1253 1303 AC repeat.  
 FT REPEAT 1318 1330 AC repeat (incomplete).  
 FT REGION 35 1050 Catalytic (approximate).  
 FT REGION 1126 1455 2.4 A, 1 C and 1 AC repeats.

REGION	1126	1455	Glucan-binding (approximate).
VARIANT	21	21	V -> I (in strain GS-5).
VARIANT	81	81	P -> L (in strain MT4239).
VARIANT	106	106	D -> V (in strain GS-5).
VARIANT	116	116	S -> A (in strain GS-5 and strain MT4467).
VARIANT	126	126	A -> T (in strain GS-5).
VARIANT	150	151	SR -> PK (in strain GS-5, strain MT4239 and strain MT4467).
VARIANT	256	256	A -> V (in strain GS-5 and strain MT4467).
VARIANT	425	425	R -> N (in strain MT4251).
VARIANT	519	519	R -> D (in strain MT4245 and strain MT4251).
VARIANT	538	538	R -> K (in strain MT4245 and strain MT4251).
VARIANT	545	545	Y -> F (in strain MT4245 and strain MT4251).
VARIANT	597	597	N -> D (in strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
VARIANT	600	600	R -> K (in strain MT4245, strain MT4251, strain MT4467 and strain MT8148).
VARIANT	601	601	A -> T (in strain GS-5).
VARIANT	614	614	M -> T (in strain GS-5).
VARIANT	727	727	T -> I (in strain MT8148).
VARIANT	734	734	A -> V (in strain MT8148).
VARIANT	964	964	L -> F (in strain MT4239).
VARIANT	1113	1113	N -> Y (in strain MT4239).
VARIANT	1118	1118	A -> T (in strain MT4239).
VARIANT	1204	1204	I -> V (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
VARIANT	1208	1208	V -> I (in strain MT8148).
VARIANT	1292	1294	DGH -> NGV (in strain GS-5, strain MT4467 and strain MT8148).
VARIANT	1305	1369	Missing (in strain MT4245).
VARIANT	1326	1326	I -> V (in strain MT8148).
VARIANT	1331	1331	T -> A (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).
VARIANT	1377	1377	R -> K (in strain MT8148).
VARIANT	1398	1398	V -> I (in strain MT8148).
VARIANT	1424	1424	D -> N (in strain MT4239).
VARIANT	1439	1439	V -> I (in strain MT4239 and strain MT8148).
VARIANT	1444	1444	S -> P (in strain MT8148).
CONFLICT	1337	1455	QSLYFKSNGVQAKGELITERKGRVYDPSNGNEVRNRYVR TSSGNWYFGNOCYALIGHVVEGRVYEDNGVYRASHD QRNWDYDYYRRDFGRGSSAVRFRHNRNGFFDNFRF -> HASILSLMVFRLRESSLQSVKVVSNMTLIPMKFVIVM (in Ref. 1).
SQ	SEQUENCE	1455 AA; 162966 MW; 3CB455A99A4FEC86 CRC64;	
Query Match	78.8%;	Score 67; DB 1; Length 1455;	
Best Local Similarity	76.9%;	Pred. No. 0.1;	
Matches	10; Conservative	2; Mismatches	1; Indels
			0; Gaps
			0;
QY	2	WNGESEKPYDDHL 14	
DB	370	WNSDSEKPFDDHL 382	
RESULT 6			
GTFFB_STRMU	STANDARD;	PRT: 1476 AA.	
ID	GTFFB_STRMU	AC P08987; O69381; O69384; O69387; O69390; O69396;	
DT	01-NOV-1988	(Rel. 09, Created)	
DT	28-FEB-2003	(Rel. 41, Last sequence update)	
DT	10-WAY-2005	(Rel. 47, Last annotation update)	
DE	Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)		
DE	(Sucrose 6-glucosyltransferase).		
GN	Name-gtfB; OrderedLocustNames=SMU.1004;		
OS	Streptococcus mutans.		
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
OC	Streptococcus.		

OX NCBI\_TaxID=1309;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=GS-5;  
 RX MEDLINE=87308013; PubMed=3040685;  
 RA Shiroza T., Ueda S., Kuramitsu H.K.;  
 RT "Sequence analysis of the *gtfB* gene from *Streptococcus mutans*.";  
 RL J. Bacteriol. 169:4263-4270(1987).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,  
 RX MT4467 / Serotype e, and MT8148 / Serotype C;  
 RL MEDLINE=98231643; PubMed=9570124;  
 RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
 RA Kimura S., Hamada S.;  
 RT "Molecular analyses of glucosyltransferase genes among strains of  
 RL *Streptococcus mutans*.";  
 RL PEMS Microbiol. Lett. 161:331-336(1998).  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=UA159 / ATCC 700610 / Serotype c;  
 RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;  
 RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
 RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,  
 RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,  
 RA Ferretti J.J.;  
 RT "Genome sequence of *Streptococcus mutans* UA159, a cariogenic dental  
 RL pathogen."; Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
 CC -1- FUNCTION: Production of extracellular glucans, that are thought to  
 play a key role in the development of the dental plaque because of  
 their ability to adhere to smooth surfaces and mediate the  
 aggregation of bacterial cells and food debris.  
 CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-  
 fructose + (1,6-alpha-D-glucosyl)(n+1).  
 CC -1- SUBCELLULAR LOCATION: Secreted  
 CC -1- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha  
 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes  
 water-soluble glucans (alpha 1,6-glucose). GTF-S synthesizes both  
 forms of glucans.  
 CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.  
 CC -1- SIMILARITY: Contains 10 cell wall binding repeats.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use as long as its content is in no way modified and this statement is not  
 removed.  
 CC -----  
 DR EMBL; M17361; AAA89588.1; -; Genomic DNA.  
 DR EMBL; D88651; BAA26101.1; -; Genomic DNA.  
 DR EMBL; D88654; BAA26105.1; -; Genomic DNA.  
 DR EMBL; D88657; BAA26109.1; -; Genomic DNA.  
 DR EMBL; D88660; BAA26113.1; -; Genomic DNA.  
 DR EMBL; D89977; BAA26119.1; -; Genomic DNA.  
 DR EMBL; AE014940; AAN58705.1; -; Genomic DNA.  
 DR PIR; B33135; B33135.  
 DR HSP; P06653; 1H8G.  
 DR InterPro; IPR002479; Cell wall bd put.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW\_binding\_1; 4.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;  
 KW Transferase.  
 FT SIGNAL 1 34 Potential.  
 FT CHAIN 35 1476 Glucosyltransferase-I.  
 FT REPEAT 1097 1130 A repeat.  
 FT REPEAT 1161 1210 1.  
 FT REPEAT 1225 1275 2.  
 FT REPEAT 1290 1340 3.  
 FT REPEAT 1355 1405 4.  
 FT REPEAT 1420 1470 5.  
 FT REGION 35 1051 Catalytic (approximate).

FT REGION 1097 1476 Glucan-binding (approximate).  
 FT REGION 1161 1470 S X tandem repeats.  
 FT VARIANT 62 S -> T (in strain MT4239).  
 FT VARIANT 65 V -> I (in strain GS-5).  
 FT VARIANT 68 V -> A (in strain GS-5, strain MT4245,  
 strain MT4251, strain MT4467 and strain  
 MT8148).  
 FT VARIANT 78 Q -> P (in strain MT4251).  
 FT VARIANT 86 I -> S (in strain GS-5, strain MT4245,  
 strain MT4251, strain MT4467 and strain  
 MT8148).  
 FT VARIANT 89 S -> F (in strain MT4251).  
 FT VARIANT 168 K -> N (in strain MT4251).  
 FT VARIANT 276 S -> D (in strain GS-5, strain MT4467 and  
 strain MT8148).  
 FT VARIANT 399 N -> R (in strain MT4239).  
 FT VARIANT 474 I -> T (in strain MT4239).  
 FT VARIANT 512 K -> R (in strain MT8148).  
 FT VARIANT 519 P -> Y (in strain MT8148).  
 FT VARIANT 701 T -> I (in strain MT8148).  
 FT VARIANT 708 A -> V (in strain MT8148).  
 FT VARIANT 938 F -> L (in strain MT8148).  
 FT VARIANT 952 FGKPE -> YGTPVA (in strain GS-5, strain  
 MT4239 and strain MT4467).  
 FT VARIANT 963 SV -> NT (in strain GS-5, strain MT4239  
 and strain MT4467).  
 FT VARIANT 968 ADS -> VDG (in strain GS-5, strain MT4239  
 and strain MT4467).  
 FT VARIANT 1086 A -> T (in strain MT4239).  
 FT VARIANT 1158 S -> N (in strain MT4239).  
 FT VARIANT 1163 H -> Y (in strain MT4251).  
 FT VARIANT 1168 E -> K (in strain MT8148).  
 FT VARIANT 1182 Y -> C (in strain MT8148).  
 FT VARIANT 1234 A -> P (in strain MT4239).  
 FT VARIANT 1263 R -> H (in strain GS-5 and strain  
 MT4467).  
 FT VARIANT 1263 R -> P (in strain MT8148).  
 FT VARIANT 1264 Y -> H (in strain GS-5, strain MT4239,  
 strain MT4467 and strain MT8148).  
 FT VARIANT 1272 S -> G (in strain GS-5, strain MT4239,  
 strain MT4467 and strain MT8148).  
 FT VARIANT 1329 H -> Y (in strain GS-5 and strain  
 MT4467).  
 FT VARIANT 1394 Y -> H (in strain GS-5, strain MT4239,  
 strain MT4467 and strain MT8148).  
 FT VARIANT 1402 S -> G (in strain GS-5, strain MT4239,  
 strain MT4467 and strain MT8148).  
 FT VARIANT 1459 Y -> H (in strain MT4467).  
 FT CONFLICT 570 R -> A (in Ref. 1).  
 FT CONFLICT 800 ADQDVRVAASTAPSTDGK -> LKMPALRLARPHQOMA  
 (in Ref. 1).  
 FT CONFLICT 1310 H -> L (in Ref. 1).  
 FT SEQUENCE 1476 AA; 165847 MW; 9C6E09F731B4CBFC CRC64;  
 SQ  
 Query Match 78.8%; Score 67; DB 1; Length 1476;  
 Best Local Similarity 76.9%; Pred. No. 0.1;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 QY 2 WNGSEKPYDDHL 14  
 Db 344 WNSDSKPFDDHL 356  
 RESULT 7  
 Q55MV2\_CRYNE PRELIMINARY; PRT; 460 AA.  
 ID Q55MV2\_CRYNE PRELIMINARY;  
 AC Q55MV2;  
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=CNBH3320;  
 OS *Cryptococcus neoformans* var. *neoformans* B-3501A.

OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;  
 OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.  
 OX NCBI\_TaxID=283643;  
 RN [1]  
 RC NUCLEOTIDE SEQUENCE.  
 RC STRAIN=B-3501A;  
 RA Fung E., Hyman R.W., Rowley D., Bruno D., Miranda M., Fukushima M.,  
 RA Wickes B.L., Fu J., Davis R.W.;  
 RA "Cryptococcus neoformans serotype D sequencing."  
 RT Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 CC EMBL; AAY01000042; EAL19233.1; -; Genomic\_DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 460 AA; 51369 MW; B8C57D2B0050AFCC CRC64;  
 Query Match 64.7%; Score 55; DB 2; Length 460;  
 Best Local Similarity 61.5%; Pred. No. 2.6;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 QWNGSEKPYDDH 13  
 Db 89 RWRGEAMKPYQDH 101  
 RESULT 8  
 ID Q5KR79\_CRYNE PRELIMINARY; PRT; 460 AA.  
 AC Q5KR79;  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
 DE Machado-Joseph disease protein 1 (Ataxin-3), putative.  
 GN ORFNames=CN103470;  
 OS Cryptococcus neoformans var. neoformans JEC21.  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;  
 OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.  
 OX NCBI\_TaxID=214684;  
 RN [1]  
 RC NUCLEOTIDE SEQUENCE.  
 RC STRAIN=JEC21;  
 RA Loftus B., Amedeo P., Roncaglia P., Vanathevan J., Utterback T.,  
 RA Van Aken S., Fraser C.;  
 RA Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=JEC21;  
 RX PubMed=15653466; DOI=10.1126/science.1103773;  
 RA Loftus B.J., Fung E., Roncaglia P., Rowley D., Amedeo P., Bruno D.,  
 RA Vanathevan J., Miranda M., Anderson I.J., Fraser J.A., Allen J.E.,  
 RA Bosdet I.E., Brent M.R., Chiu R., Doering T.L., Donlin M.J.,  
 RA D'Souza C.A., Fox D.S., Grinberg V., Fu J., Fukushima M., Haas B.J.,  
 RA Huang J.C., Janbon G., Jones S.J.M., Koo H.L., Krzywinski M.I.,  
 RA Kwon-Chung K.J., Lengell K.B., Maiti R., Marra M.A., Marra R.E.,  
 RA Mathewson C.A., Mitchell T.G., Perlea M., Riggs F.R., Salzberg S.L.,  
 RA Schein J.E., Shvartsbeyn A., Shin H., Shumway M., Specht C.A.,  
 RA Sun B.B., Tenney A., Utterback T.R., Wickes B.L., Wortman J.R.,  
 RA Wye N.H., Kronstad J.W., Lodge J.K., Heitman J., Davis R.W.,  
 RA Fraser C.M., Hyman R.W.;  
 RA "The genome of the basidiomycetous yeast and human pathogen  
 RT Cryptococcus neoformans."  
 RL Science 307.1329-1324(2005).  
 DR EMBL; AE017349; AAW45308.1; -; Genomic\_DNA.  
 DR InterPro; IPR006155; Josephin.  
 DR Pfam; PF02099; Josephin; 1.  
 DR PRINTS; PR01233; JOSEPHIN.  
 DR PROSITE; PS50957; JOSEPHIN; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 460 AA; 51369 MW; B8C57D2B0050AFCC CRC64;  
 Query Match 64.7%; Score 55; DB 2; Length 460;  
 Best Local Similarity 61.5%; Pred. No. 2.6;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 QWNGSEKPYDDH 13  
 Db 89 RWRGEAMKPYQDH 101  
 RESULT 9  
 ID Q84CN4\_LEUME PRELIMINARY; PRT; 1330 AA.  
 AC Q84CN4;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Dextranucrase DsrR (EC 2.4.1.5).  
 GN Name=dsrR;  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=NRRL B-1501;  
 RA Kim C.H., Moon J.O., Jang E.K.;  
 RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AV142210; AAN38835.1; -; Genomic DNA.  
 DR GO; GO:0047849; F:dextranucrase activity; IEA.  
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW binding\_1; 1.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Glycosyltransferase; Transferase.  
 SQ SEQUENCE 1330 AA; 148863 MW; D945CBB36CF75797 CRC64;  
 Query Match 60.0%; Score 51; DB 2; Length 1330;  
 Best Local Similarity 64.3%; Pred. No. 39;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 QWNGSEKPYDDH 14  
 Db 250 QWNMSDEPKNDHL 263  
 RESULT 10  
 ID Q9L466\_LEUME PRELIMINARY; PRT; 1477 AA.  
 AC Q9L466;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Dextranucrase (EC 2.4.1.5).  
 GN Name=dsrC;  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=NRRL B-1355;  
 RA Arguello-Morales M.A., Remaud-Simeon M., Pizzut S., Sarcabal P.,  
 RA Willmot R.M., Monsan P.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ250172; CAB76565.1; -; Genomic DNA.  
 DR GO; GO:0047849; F:dextranucrase activity; IEA.  
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW binding\_1; 1.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 KW Glycosyltransferase; Transferase.  
 SQ SEQUENCE 1477 AA; 164886 MW; E6F5710DEDFCB831 CRC64;

Query Match 60.0%; Score 51; DB 2; Length 1477;  
 Best Local Similarity 64.3%; Pred. No. 44;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDHL 14  
 ||| ||| ||| :|||  
 DB 397 QWNMSSEDPKNDHL 410

RESULT 11

ID Q3EZH5\_LEUME PRELIMINARY; PRT; 1508 AA.  
 AC Q3EZH5;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Dextranucrase Derb742.  
 GN Name=derb742;  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=B-742CB;  
 RA Kim H.-S., Kim D., Ryu H.-J., Robyt J.F.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF294469; AAG38021.1; -; Genomic DNA.  
 DR GO; GO:0009250; P:Glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; Glyco\_hydro\_70.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR KW Glycosyltransferase; Transferrase.  
 DR SQ SEQUENCE 1508 AA; 168542 MW; E2FCFA0F87AE4F3A CRC64;

Query Match 60.0%; Score 51; DB 2; Length 1508;  
 Best Local Similarity 64.3%; Pred. No. 45;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDHL 14  
 ||| ||| ||| :|||  
 DB 428 QWNMSSEDPKNDHL 441

RESULT 12

ID O52224\_LEUME PRELIMINARY; PRT; 1508 AA.  
 AC O52224;  
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Glucosyltransferase (EC 2.4.1.5).  
 GN Name=d8rB;  
 OS Leuconostoc mesenteroides.  
 OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
 OX NCBI\_TaxID=1245;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=NRRL B-1299;  
 RA Monchois V., Renaud-Simeon M., Monsan P., Willemot R.M.;  
 RT "Cloning and sequencing of a gene coding for an extracellular  
 RT dextranucrase (DSRB) from Leuconostoc mesenteroides NRRL B-1299  
 RT synthesizing only a (1-6) glucan.";  
 RL FEMS Microbiol. Lett. 0:0-0(1998).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed=9503626;  
 RA Monchois V., Renaud-Simeon M., Monsan P., Willemot R.M.;  
 RT "Cloning and sequencing of a gene coding for an extracellular  
 RT dextranucrase (DSRB) from Leuconostoc mesenteroides NRRL B-1299  
 RT synthesizing only a (1-6) glucan.";  
 RL FEMS Microbiol. Lett. 159:307-315(1998).  
 DR EMBL; AF030129; AAB95453.1; -; Genomic\_DNA.  
 DR PIR; T31098; T31098.

DR GO; GO:0047849; F:dextranucrase activity; IEA.  
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR Pfam; PF01473; CW binding\_1; 1.  
 DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
 DR KW Glycosyltransferase; Transferrase.  
 DR SQ SEQUENCE 1508 AA; 168511 MW; E70CECB57A70D1F0 CRC64;

Query Match 60.0%; Score 51; DB 2; Length 1508;  
 Best Local Similarity 64.3%; Pred. No. 45;  
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDHL 14  
 ||| ||| ||| :|||  
 DB 428 QWNMSSEDPKNDHL 441

RESULT 13

ID METJ\_PHOLL STANDARD; PRT; 105 AA.  
 AC O7MYC7;  
 DT 29-MAR-2004 (Rel. 43, Created)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 13-SEP-2005 (Rel. 48, Last annotation update)  
 DE Met repressor (Met regulon regulatory protein metJ).  
 GN Name=metJ; OrderedLocNames=plu4757;  
 OS Photorhabdus luminescens (subsp. laumondii).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Photorhabdus.  
 OX NCBI\_TaxID=141679;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=TT01;  
 RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;  
 RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,  
 RA Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,  
 RA Dassa E., Deroose R., Derzelle S., Freyassinet G., Gaudriault S.,  
 RA Medigue C., Lanois A., Powell K., Siguer P., Vincent R., Wingate V.,  
 RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;  
 RT "The genome sequence of the entomopathogenic bacterium Photorhabdus  
 RT luminescens.";  
 RL Nat. Biotechnol. 21:1307-1313(2003).  
 CC -!- FUNCTION: This regulatory protein, when combined with SAM (S-  
 CC adenosylmethionine) represses the expression of the methionine  
 CC regulon and of enzymes involved in SAM synthesis (By similarity).  
 CC -!- SUBUNIT: Homodimer (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -!- DOMAIN: Does not bind DNA by an helix-turn-helix motif.  
 CC -!- SIMILARITY: Belongs to the metJ family.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use as long as its content is in no way modified and this statement is not  
 CC removed.  
 CC -----  
 DR EMBL; BX571874; CAEL17129.1; -; Genomic\_DNA.  
 DR SMR; O7MYC7; 2-105.  
 DR Photolist; plu4757; -.  
 DR HAMAP; MF\_00744; -; 1.  
 DR InterPro; IPR002084; MetJ.  
 DR Pfam; PF01340; MetJ; 1.  
 DR PIRSF; PIRSF003191; MetJ; 1.  
 DR ProDom; PD020365; MetJ; 1.  
 DR ProDom; PD020365; MetJ; 1.  
 KW Amino-acid biosynthesis; Complete proteome; DNA-binding;  
 KW Methionine biosynthesis; Repressor; Transcription;  
 KW Transcription regulation.  
 DR SQ SEQUENCE 105 AA; 12235 MW; 926DEDB5320BC8F5 CRC64;

Query Match 56.5%; Score 48; DB 1; Length 105;  
 Best Local Similarity 53.8%; Pred. No. 6.9;

```
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 1 QWNGESEKPYDDH 13
Db 3 EWNGEYVSPAHE 15

RESULT 14
METJ_YERPE
ID METJ_YERPE STANDARD; PRT; 105 AA.
AC Q8ZJ18;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Met repressor (Met regulon regulatory protein metJ).
GN Name=metJ; OrderedLocusNames=IP00114, Y0301, YP0116;
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CO-92 / Biovar Orientalis;
RA MEDLINE=21170413; PubMed=11586360; DOI=10.1038/35097083;
RA Parkhill J., Wren B.W., Thomson N.R., Tittball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C.M., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.-M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holtroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.M.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=KIMS / Biovar Mediaevalis;
RA MEDLINE=22137863; PubMed=12142430;
RX DOI=10.1128/JB.184.16.4601-4611.2002;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=91001 / Biovar Mediaevalis;
RX PubMed=15368893;
RA Song Y., Tong Z., Wang J., Wang L., Guo Z., Han Y., Zhang J., Pei D.,
RA Zhou D., Qin H., Pang X., Han Y., Zhai J., Li M., Cui B., Qi Z.,
RA Jin L., Dai R., Chen F., Li S., Ye C., Du Z., Lin W., Wang J., Yu J.,
RA Yang H., Wang J., Huang P., Yang R.;
RT "Complete genome sequence of Yersinia pestis strain 91001, an isolate
avirulent to humans.";
RL DNA Res. 11:179-197(2004).
CC -!- FUNCTION: This regulatory protein, when combined with SAM (S-
adenosylmethionine) represses the expression of the methionine
regulon and of enzymes involved in SAM synthesis (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- DOMAIN: Does not bind DNA by an helix-turn-helix motif.
CC -!- SIMILARITY: Belongs to the metJ family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; AJ414141; CAC88978.1; -; Genomic DNA.
CC EMBL; AB013629; AAM83893.1; -; Genomic DNA.
CC EMBL; AB017127; AAS60395.1; -; Genomic DNA.
```

```
DR PIR; AH0014; AH0014.
DR HSP; P08338; 1CMB.
DR SMR; Q8ZJ18; 2-105.
DR HMAP; MF_00744; -, 1.
DR InterPro; IPR002084; MetJ.
DR Pfam; PF01340; MetJ; 1.
DR PIRSF; PIRSF003191; MetJ; 1.
DR ProDom; PD020365; MetJ; 1.
KW Amino-acid biosynthesis; Complete proteome; DNA-binding;
KW Methionine biosynthesis; Repressor; Transcription;
KW Transcription regulation.
SQ SEQUENCE 105 AA; 12149 MW; 59984C46320582BF CRC64;

Query Match 56.5%; Score 48; DB 1; Length 105;
Best Local Similarity 53.8%; Pred. No. 6.9;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDH 13
Db 3 EWNGEYVSPAHE 15

RESULT 15
Q6G78_YERPS
ID Q6G78_YERPS PRELIMINARY; PRT; 105 AA.
AC Q6G78;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Transcriptional repressor protein.
GN Name=metJ; OrderedLocusNames=YPTB0104;
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=633;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=IP32953 / Serotype I;
RX PubMed=15358858; DOI=10.1073/pnas.0404012101;
RA Chain P.S.G., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O.,
RA Regala W.M., Georgescu A.M., Vergez L.M., Land M.L., Motin V.L.,
RA Brubaker R.R., Fowler J., Hinnébusch J., Marceau M., Medigue C.,
RA Simonet M., Chenal-Francois V., Souza B., Dacheux D., Elliott J.M.,
RA Derbise A., Hauser L.J., Garcia E.;
RT "Insights into the evolution of Yersinia pestis through whole-genome
comparison with Yersinia pseudotuberculosis.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:13826-13831(2004).
DR EMBL; BX936398; CAH19344.1; -; Genomic DNA.
DR SMR; Q6G78; 2-105.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0003086; P:methionine biosynthesis; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR002084; MetJ.
DR Pfam; PF01340; MetJ; 1.
DR ProDom; PD020365; MetJ; 1.
KW Complete proteome.
SQ SEQUENCE 105 AA; 12150 MW; B97642A8320F881F CRC64;

Query Match 56.5%; Score 48; DB 2; Length 105;
Best Local Similarity 53.8%; Pred. No. 6.9;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDH 13
Db 3 EWNGEYVSPAHE 15

RESULT 16
Q6CZA0_ERWCT
ID Q6CZA0_ERWCT PRELIMINARY; PRT; 105 AA.
AC Q6CZA0;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
```

25-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Repressor of the methionine regulon.  
 GN Name=metJ; OrderedLocNames=ECA4253;  
 OS *Erwinia carotovora* (subsp. atroseptica) (Pectobacterium atrosepticum).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Pectobacterium.  
 OX NCBI\_TaxID=29471;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=SCRI 1043 / ATCC BRA-672;  
 RX PubMed=15263089; DOI=10.1073/pnas.0402424101;  
 RA Bell K.S., Sebahia M., Pritchard L., Holden M.T.G., Hyman L.J.,  
 RA Holleva M.C., Thomson N.R., Bentley S.D., Churcher L.J.C., Mungall K.,  
 RA Atkin R., Bason N., Brooks K., Chillingworth T., Clark K., Doggett J.,  
 RA Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H.,  
 RA Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S.,  
 RA Salmond G.P.C., Birch P.R.J., Parkhill J., Toth I.K.;  
 RT "Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. atroseptica and characterization of virulence factors."  
 RT factors."  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:11105-11110(2004).  
 DR EMBL; BX950851; CAG77150.1; -, Genomic\_DNA.  
 DR SMR; O6CZAO; 2-105.  
 DR GO; GO:0003700; P:transcription factor activity; IEA.  
 DR GO; GO:0009086; P:methionine biosynthesis; IEA.  
 DR GO; GO:006355; P:regulation of transcription, DNA-dependent; IEA.  
 DR InterPro; IPR002084; MetJ.  
 DR Pfam; PF01340; MetJ; 1.  
 DR PIRSF; PIRSF003191; MetJ; 1.  
 DR ProDom; PD020365; MetJ; 1.  
 DR Complete proteome.  
 KW SEQUENCE 105 AA; 12182 MW; 7A7707A8320BD90D CRC64;  
 SQ

Query Match 56.5%; Score 48; DB 2; Length 105;  
 Best Local Similarity 53.8%; Pred. No. 6.9;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QWGESEKPYDDH 13  
 :||||| :| :|  
 Db 3 EWNGEYVSPYAEH 15

RESULT 17  
 Q4TTA0 TETNG  
 ID Q4TTA0 TETNG PRELIMINARY; PRT; 262 AA.  
 AC Q4TTA0;  
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Chromosome undetermined SCAF2302, whole genome shotgun sequence.  
 DE (Fragment).  
 GN ORFNames=GSTENG00038394001;  
 OS Tetraodon nigroviridis (Green puffer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraontoidea; Tetraodontidae; Tetraodon.  
 OX NCBI\_TaxID=99883;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RG Genoscope; Whitehead Institute Centre for Genome Research;  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC preliminary data.  
 DR EMBL; CAAG01014735; CAG04191.1; -, Genomic\_DNA.  
 FT NON TER 1 1  
 SQ SEQUENCE 663 AA; 75795 MW; A51C2C31C01F6D72 CRC64;

Query Match 56.5%; Score 48; DB 2; Length 663;  
 Best Local Similarity 58.3%; Pred. No. 55;  
 Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 WNGESEKPYDDH 13  
 :||||| :| :|  
 Db 554 WQGERHHPYDQH 565

RESULT 18  
 Q4S591 TETNG  
 ID Q4S591 TETNG PRELIMINARY; PRT; 663 AA.  
 AC Q4S591;  
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Chromosome undetermined SCAF14735, whole genome shotgun sequence.  
 DE (Fragment).  
 GN ORFNames=GSTENG00023836001;  
 OS Tetraodon nigroviridis (Green puffer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraontoidea; Tetraodontidae; Tetraodon.  
 OX NCBI\_TaxID=99883;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RG Genoscope; Whitehead Institute Centre for Genome Research;  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC preliminary data.  
 DR EMBL; CAAG01014735; CAG04191.1; -, Genomic\_DNA.  
 FT NON TER 1 1  
 SQ SEQUENCE 663 AA; 75795 MW; A51C2C31C01F6D72 CRC64;

Query Match 56.5%; Score 48; DB 2; Length 663;  
 Best Local Similarity 58.3%; Pred. No. 55;  
 Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 WNGESEKPYDDH 13  
 :||||| :| :|  
 Db 554 WQGERHHPYDQH 565

RT the early vertebrate proto-karyotype.";  
 RL Nature 431:946-957(2004).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RG Genoscope; Whitehead Institute Centre for Genome Research;  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC preliminary data.  
 DR EMBL; CAAG01002302; CAF87382.1; -, Genomic\_DNA.  
 FT NON TER 1 1  
 SQ SEQUENCE 262 AA; 30639 MW; 7BC558571F040AAB CRC64;

Query Match 56.5%; Score 48; DB 2; Length 262;  
 Best Local Similarity 58.3%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 WNGESEKPYDDH 13  
 :||||| :| :|  
 Db 152 WQGERHHPYDQH 163

RESULT 18  
 Q4S591 TETNG  
 ID Q4S591 TETNG PRELIMINARY; PRT; 663 AA.  
 AC Q4S591;  
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Chromosome undetermined SCAF14735, whole genome shotgun sequence.  
 DE (Fragment).  
 GN ORFNames=GSTENG00023836001;  
 OS Tetraodon nigroviridis (Green puffer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraontoidea; Tetraodontidae; Tetraodon.  
 OX NCBI\_TaxID=99883;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RG Genoscope; Whitehead Institute Centre for Genome Research;  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC preliminary data.  
 DR EMBL; CAAG01014735; CAG04191.1; -, Genomic\_DNA.  
 FT NON TER 1 1  
 SQ SEQUENCE 663 AA; 75795 MW; A51C2C31C01F6D72 CRC64;

Query Match 56.5%; Score 48; DB 2; Length 663;  
 Best Local Similarity 58.3%; Pred. No. 55;  
 Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 WNGESEKPYDDH 13  
 :||||| :| :|  
 Db 554 WQGERHHPYDQH 565

```

RESULT 19
Q4T6C1_TETNG
ID Q4T6C1_TETNG PRELIMINARY; PRT; 754 AA.
AC Q4T6C1;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Chromosome undetermined SCAP8828, whole genome shotgun sequence.
DE (Fragment).
GN ORFNames=GSTENG0006388001;
OS Tetraodon nigroviridis (Green puffer).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
OC Tetraodontidae; Tetraodontidae; Tetraodon.
OX NCBI_TaxID=99883;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Jaillon O., Aury J.M., Brunet F., Petit J.L., Stange-Thomann N.,
RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,
RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,
RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,
RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,
RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,
RA Cruaud C., Duprat S., Brottier P., Coutanceau J.P., Gouzy J.,
RA Parra G., Lardier G., Chappie C., McKernan K.J., McEwan P., Bosak S.,
RA Kellis M., Wolff J.N., Guigo R., Zody M.C., Mesirov J.,
RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,
RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,
RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;
RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals
RT the early vertebrate proto-karyotype.";
RL Nature 431:946-957(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Genoscope; Whitehead Institute Centre for Genome Research;
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; CAABE01008828; CAP91561.1; -; Genomic_DNA.
FT NON TER 1
FT SEQUENCE 754 AA; 86915 MW; FE4A7EEB1646CAAB CRC64;
SQ
Query Match 56.5%; Score 48; DB 2; Length 754;
Best Local Similarity 58.3%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 WNGESEKPYDDH 13
DB 644 WQGERHHPYDOH 655

RESULT 20
Q6BGG9_PARTE
ID Q6BGG9_PARTE PRELIMINARY; PRT; 1271 AA.
AC Q6BGG9;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Multispecific organic anion transporter, putative.
GN ORFNames=PTMB.54c;
OS Paramecium tetraurelia.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Peniculida;
OC Paramecium.
OX NCBI_TaxID=5888;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA STRAIN=Stock d4-2;
RX PubMed=15296759; DOI=10.1016/j.cub.2004.07.029;
RA Zagulski M., Nowak J.K., Le Mouel A., Nowacki M., Migdalski A.,
Gromadka R., Noel B., Blanc I., Dessen P., Wincker P., Keller A.M.,
Cohen J., Meyer E., Sperling L.;
RT "High Coding Density on the Largest Paramecium tetraurelia Somatic
RT Chromosome.";
RL Curr. Biol. 14:1397-1404(2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA STRAIN=Stock d4-2;
RA Nowak J.K., Migdalski A., Gromadka R., Zagulski M.;
RT "Paramecium megabase sequencing project.";
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; CR548612; CAH03251.1; -; Genomic DNA.
DR GO; GO:0016021; F:ATP binding; IEA.
DR GO; GO:0005524; F:ATPase activity; IEA.
DR GO; GO:0016887; F:ATPase activity; IEA.
DR GO; GO:0042826; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR011527; ABC_membrane_1.
DR InterPro; IPR001140; ABC_TM_transp.
DR InterPro; IPR003439; ABC_transp_like.
DR Pfam; PF00664; ABC_membrane_2.
DR Pfam; PF00005; ABC_tran; 2.
DR ProDom; PD000006; ABC_transporter; 2.
DR SMART; SM00382; AAA; 2.
DR PROSITE; PS50929; ABC_TM1F; 2.
DR PROSITE; PS00211; ABC_TRANSPORTER_1; 2.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 2.
SQ SEQUENCE 1271 AA; 145387 MW; FC256F257F68F55F CRC64;

Query Match 56.5%; Score 48; DB 2; Length 1271;
Best Local Similarity 50.0%; Pred. No. 11e+02;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDHL 14
DB 737 QWNGQDQLTYDDNL 750

RESULT 21
METJ_SHEON
ID METJ_SHEON STANDARD; PRT; 104 AA.
AC QBEA52;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Met repressor (Met regulon regulatory protein metJ).
GN Name=metJ; OrderedLocustNames=SQ4057;
OS Shewanella oneidensis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Shewanellaceae; Shewanella.
OX NCBI_TaxID=70863;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=NR-1;
RX MEDLINE=22297686; PubMed=12368813; DOI=10.1038/nbt749;
RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
RA Read T.D., Eisen J.A., Seshadri R., Ward N.L., Methe B.A.,
RA Clayton R.A., Meyer T., Tsapin A., Scott J., Beanan M.J.,
RA Brinkac L.M., Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S.,
RA Haft D.H., Kolonay J.F., Madupu R., Peterson J.D., Umayam L.A.,
RA White O., Wolf A.M., Vamathevan J.J., Weidman J.F., Impraim M.,
RA Lee K., Berry K.J., Lee C., Mueller J., Khouri H.M., Gill J.,
RA Uterback T.R., McDonald L.A., Feldblyum T.V., Smith H.O.,
RA Venter J.C., Nealon K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis.";
RL Biotechnol. 20:1118-1123(2002).
CC -!- FUNCTION: This regulatory protein, when combined with SAM (S-
CC adenosylmethionine) represses the expression of the methionine
CC regulon and of enzymes involved in SAM synthesis (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).

```



```

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- DOMAIN: Does not bind DNA by an helix-turn-helix motif.
CC -1- SIMILARITY: Belongs to the metJ family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; AE015837; AAN57031.1; -; Genomic_DNA.
DR HSSP; P08338; 1CMB.
DR SMR; Q8EA52; 3-104.
DR TIGR; SO4057; -.
DR HAMAP; MF_00744; -; 1.
DR InterPro; IPR002084; MetJ.
DR Pfam; PF01340; MetJ; 1.
DR PIRSF; PIRSF003191; MetJ; 1.
DR ProDom; PD020365; MetJ; 1.
KW Amino-acid biosynthesis; Complete proteome; DNA-binding;
KW Methionine biosynthesis; Repressor; Transcription;
KW Transcription regulation.
SQ SEQUENCE 104 AA; 11959 MW; 9FE7A8DB33C1A9CB CRC64;

Query Match 55.3%; Score 47; DB 1; Length 104;
Best Local Similarity 53.8%; Pred. No. 10;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDH 13
DB 3 EWNGEYISPYAEH 15

RESULT 22
Q6LVH6 PHOPR
ID Q6LVH6 PHOPR PRELIMINARY; PRT; 105 AA.
AC Q6LVH6
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative met repressor.
GN OrderedLocustNames=BPRA0260;
OS Photobacterium profundum (Photobacterium sp. (strain SS9)).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Photobacterium.
OX NCBI_TaxID=74109;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=15746425; DOI=10.1126/science.1103341;
RA Vezzi A., Campanaro S., D'Angelo M., Simonato F., Vitulo N.,
RA Lauro F.M., Cestaro A., Malacrida G., Simonati B., Cannata N.,
RA Romualdi C., Bartlett D.H., Valle G.;
RT "Life at depth: Photobacterium profundum genome sequence and
RT expression analysis.";
RL Science 307:1459-1461(2005).
DR EMBL; CR378663; CAG18699.1; -; Genomic_DNA.
DR SMR; Q6LVH6; 2-101.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0009086; P:methionine biosynthesis; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR002084; MetJ.
DR Pfam; PF01340; MetJ; 1.
DR PIRSF; PIRSF003191; MetJ; 1.
DR ProDom; PD020365; MetJ; 1.
KW Complete proteome.
SQ SEQUENCE 105 AA; 11971 MW; 43CBA978C8D5CFBB CRC64;

Query Match 55.3%; Score 47; DB 2; Length 105;
Best Local Similarity 53.8%; Pred. No. 10;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDH 13
DB 3 EWNGEYISPYAEH 15

```

```

Db 3 EWNGEYISPYAEH 15

RESULT 23
Q5SBM6 LACFE
ID Q5SBM6 LACFE PRELIMINARY; PRT; 1463 AA.
AC Q5SBM6
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Glucansucrase (EC 2.4.1.5).
OS Lactobacillus fermentum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1613;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Kg3;
RX PubMed=15528655; DOI=10.1099/mic.0.27321-0;
RA Kralj S., van Geel-Schutten G.H., Dondorff M.M.G., Kirsanovs S.,
RA van der Maarel M.J.E.C., Dijkhuizen L.;
RT "Glucan synthesis in the genus Lactobacillus: isolation and
RT characterization of glucansucrase genes, enzymes and glucan products
RT from six different strains.";
RL Microbiology 150:3681-3690(2004).
DR EMBL; AY697433; AAU08008.1; -; Genomic_DNA.
DR GO; GO:0047849; F:dextranase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR003138; Glyco_hydro_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1463 AA; 161049 MW; F7EAB56AB1A3A68 CRC64;

Query Match 55.3%; Score 47; DB 2; Length 1463;
Best Local Similarity 64.3%; Pred. No. 2e+02;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDH 14
DB 468 QWNKTSVDNDH 481

RESULT 24
Q4QJG4 LEIMA
ID Q4QJG4 LEIMA PRELIMINARY; PRT; 1819 AA.
AC Q4QJG4
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Microtubule-associated protein, putative.
GN ORFNames=LmjF05_0380;
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Friedlin;
RA Peacock C.S., Murphy L., Ivens A.C., Berriman M., Blackwell J.,
RA Smith D., Collins M., Foster N., Harris D., Oliver K., O'Neil S.,
RA Saunders D., Seeger K., Warren T., Apostolou Z., Bauser C., Beck A.,
RA Blanchette G., Borzym K., Bothe G., Bruschi C., Ciaroni L.,
RA Duesterhoeft A., Fuchs M., Gabel C., Goffeau A., Hilbert H.,
RA Klages S., Kube M., Marra M., Masuy D., Mueller-Auer S., Pohl T.,
RA Kurnelle B., Reinhardt R., Rieger M., Robben J., Schaefer M.,
RA Tosato V., Volckaert G., Wambutt R., Wedler H., Zimmermann M.,
RA Rajandream M., Barrell B.G.;
RL Submitted (JUN-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; CT005244; CAJ01958.1; -; Genomic_DNA.
SQ SEQUENCE 1819 AA; 206840 MW; 7DF2CC4B721BB5BF CRC64;

Query Match 55.3%; Score 47; DB 2; Length 1819;
Best Local Similarity 61.5%; Pred. No. 2.5e+02;

```

Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 QWNGSEKPYDDH 13  
 |||: |||  
 Db 29 QWNGLEKPYDDH 41

RESULT 25  
 Q8IBX2\_PLAF7  
 ID Q8IBX2\_PLAF7 PRELIMINARY; PRT; 2275 AA.  
 AC Q8IBX2;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Erythrocyte membrane protein 1 (PfEMP1).  
 GN Name=PF010049; Synonyms=VAR;  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]

RP NUCLEOTIDE SEQUENCE.  
 RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,  
 RA Quail M., Barrell B.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL844506; CAD50870.1; -; Genomic DNA.  
 DR GO; GO:0005539; F:glycosaminoglycan binding; IEA.  
 DR GO; GO:0009405; P:pathogenesis; IEA.  
 DR InterPro; IPR004258; PfEMP.  
 DR Pfam; PF03011; PfEMP; 2.  
 SQ SEQUENCE 2275 AA; 260958 MW; E4A56BEDF42D1F9C CRC64;

Query Match 55.3%; Score 47; DB 2; Length 2275;  
 Best Local Similarity 61.5%; Pred. No. 3.2e+02;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 QWNGSEKPYDDH 13  
 |||: |||  
 Db 880 QWNGEQPVDDH 892

RESULT 26  
 Q56X20\_ARATH  
 ID Q56X20\_ARATH PRELIMINARY; PRT; 166 AA.  
 AC Q56X20;  
 DT 10-MAY-2005 (TrEMBLrel. 30, Created)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
 DE Hypothetical protein At5g22350 (Fragment).  
 GN Name=At5g22350;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]

RP NUCLEOTIDE SEQUENCE.  
 RA Totoki Y., Seki M., Ishida J., Nakajima M., Enju A., Kamiya A.,  
 RA Narusaka M., Shin-i T., Nakagawa M., Sakamoto N., Oishi K., Kohara Y.,  
 RA Kobayashi M., Toyoda A., Sakaki Y., Sakurai T., Iida K., Akiyama K.,  
 RA Satou M., Toyoda T., Konegaya A., Carninci P., Kawai J.,  
 RA Hayashizaki Y., Shinozaki K.;  
 RT "Large-scale analysis of RIKEN Arabidopsis full-length (RAFL) cDNAs."  
 RL Submitted (MAR-2005) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK221859; BAD94137.1; -; mRNA.  
 KW Hypothetical protein.  
 FT NON TER  
 SQ SEQUENCE 166 AA; 18476 MW; 44E10CALA0021397 CRC64;

Query Match 54.1%; Score 46; DB 2; Length 166;  
 Best Local Similarity 53.8%; Pred. No. 25;  
 Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDH 14

Db 62 WNGQEPNPGHL 74  
 |||: |||

RESULT 27  
 Q4WTQ5\_ASPPFU  
 ID Q4WTQ5\_ASPPFU PRELIMINARY; PRT; 424 AA.  
 AC Q4WTQ5;  
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=Afu5g05730;  
 OS Aspergillus fumigatus Af293.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=330879;  
 RN [1]

RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=Af293;  
 RA Nierman W., Pain A., Anderson M.J., Wortman J., Kim H.Stanley.,  
 RA Arroya J., Berriman M., Abe K., Archer D.B., Bernerjo C., Bennett J.,  
 RA Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,  
 RA Farman M., Fedorova N., Fedorova N., Feldblyum T.V., Fischer R.,  
 RA Fosker N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,  
 RA Goldman G.H., Gomi K., Griffith-Jones S., Gwilliam R., Haas B.,  
 RA Haas H., Harris D., Horiuchi H., Huang J., Humphrey S., Jimenez J.,  
 RA Keller N., Khouri H., Kitamoto K., Kobayashi T., Kulkarni R.,  
 RA Kumagai T., Laiton A., Latge J.-P., Li W., Lord A., Lu C.,  
 RA Majoros W.H., May G.S., Miller B.L., Mohamoud Y., Molina M., Monod M.,  
 RA Mounya I., Mulligan S., Murphy L., O'Neil S., Paulsen I.,  
 RA Penava M.A., Pertea M., Price C., Pritchard B.L., Quail M.A.,  
 RA Rabinowitz E., Rawlins N., Rajandream M.-A., Reichard U.,  
 RA Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,  
 RA Renning C.M., Rutter S., Salzberg S.L., Sanchez M.,  
 RA Takeuchi M., Tekala F., Saunders D., Seeger K., Squares S.,  
 RA Takeuchi M., Tekala F., Turner G., Vazquez de Aldana C.R., Weidman J.,  
 RA White O., Woodward J., Yu J.-H., Fraser C., Galagan J.E., Asai K.,  
 RA Machida M., Hall N., Barrell B., Denning D.W.;  
 RT "Genomic sequence of the pathogenic and allergenic filamentous fungus  
 Aspergillus fumigatus.";  
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AAHF01000003; EAL92021.1; -; Genomic\_DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 424 AA; 47773 MW; CB5B3A8B62CD9457 CRC64;

Query Match 54.1%; Score 46; DB 2; Length 424;  
 Best Local Similarity 58.3%; Pred. No. 71;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDH 13  
 |||: |||  
 Db 268 WNGDEDEDDH 279

RESULT 28  
 Q93YN4\_ARATH  
 ID Q93YN4\_ARATH PRELIMINARY; PRT; 427 AA.  
 AC Q93YN4;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)  
 DE Hypothetical protein MWD9.14 (Hypothetical protein At5g22350).  
 GN Name=MWD9.14; Synonyms=At5g22350;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 OX NCBI\_TaxID=3702;  
 RN [1]

```

RP NUCLEOTIDE SEQUENCE.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RA Tripp M., Southwick A., Karlin-Neumann G., Nguyen M., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR ENBL; AY059915; AAL24397.1; -; mRNA.
DR ENBL; AY128808; AAM91208.1; -; mRNA.
DR InterPro; IPR009367; DUF1022.
DR Pfam; PF06258; DUF1022; 1.
KW Hypothetical protein.
SQ SEQUENCE 427 AA; 47604 MW; 3A81A179A1BF21C6 CRC64;

Query Match 54.1%; Score 46; DB 2; Length 427;
Best Local Similarity 53.8%; Pred. No. 71;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 2 WNGSEKPYDDHL 14
Db 323 WNGQEPNPMGHL 335

RESULT 29
Q9FMF7 ARATH PRELIMINARY; PRT; 491 AA.
AC Q9FMF7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Arabidopsis thaliana genomic DNA, chromosome 5, P1 clone:MWD9.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=98162728; PubMed=9501997;
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned P1 clones."
RL DNA Res. 4:401-414(1997).
DR ENBL; AB007651; BAB08333.1; -; Genomic_DNA.
DR InterPro; IPR009367; DUF1022.
DR Pfam; PF06258; DUF1022; 1.
SQ SEQUENCE 491 AA; 54898 MW; D2273A90E39F2B7A CRC64;

Query Match 54.1%; Score 46; DB 2; Length 491;
Best Local Similarity 53.8%; Pred. No. 84;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 2 WNGSEKPYDDHL 14
Db 387 WNGQEPNPMGHL 399

RESULT 30
O97296_PLAF7
ID O97296_PLAF7 PRELIMINARY; PRT; 899 AA.
AC O97296;

```

```

DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein MAL3P7.30.
GN Names=MAL3P7.30; Synonyms=PF0C1000w;
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=99376085; PubMed=10448855; DOI=10.1038/22964;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Gentles P., Gwilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagels K., Jaseal K., Kyes S., McLean J., Moule S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Bartell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
RT falciparum."
RL Nature 400:532-538 (1999).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22255708; PubMed=12368867; DOI=10.1038/nature01095;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Akin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinovitch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Bartell B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."
RL Nature 419:527-531(2002).
DR ENBL; AL034559; CAB39043.2; -; Genomic_DNA.
DR InterPro; IPR009772; D123.
DR Pfam; PF07065; D123; 1.
KW Hypothetical protein.
SQ SEQUENCE 899 AA; 108575 MW; 161CE83F3B0E9613 CRC64;

Query Match 54.1%; Score 46; DB 2; Length 899;
Best Local Similarity 46.2%; Pred. No. 1.7e+02;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 QWNGSEKPYDDH 13
Db 201 KWNGDIDSIFDDH 213

RESULT 31
METJ_VIBCH
ID METJ_VIBCH STANDARD; PRT; 105 AA.
AC Q9KNF9;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Met repressor (Met regulon regulator protein metJ).
GN Name=metJ; OrderedLocusNames=VC2682;
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301; DOI=10.1038/35020000;

```

```

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.L.,
RA Emolaeva M.D., Vamathevan J.J., Baas S., Qin H., Dragol I.,
RA Sellers P., McDonald L.A., Uytterback T.R., Fleischmann R.D.,
RA Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R.,
RA Mekalanos J.J., Venter J.C., Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483 (2000).
CC -!- FUNCTION: This regulatory protein, when combined with SAM (S-
CC adenosylmethionine) represses the expression of the methionine
CC regulon and of enzymes involved in SAM synthesis (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- DOMAIN: Does not bind DNA by an helix-turn-helix motif.
CC -!- SIMILARITY: Belongs to the metJ family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC ENBL; AE004333; AAF95823.1; -; Genomic_DNA.
CC HSP; E82047; E82047.
CC HSP; P08338; 1CMB.
CC SMR; Q9KNP9; 2-104.
CC TIGR; VC2682; -.
CC HAMAP; MF 00744; -; 1.
CC InterPro; IPR002084; MetJ.
CC Pfam; PF01340; MetJ; 1.
CC PIRSF; PIRSF003191; MetJ; 1.
CC ProDom; PD020365; MetJ; 1.
CC Amino-acid biosynthesis; Complete proteome; DNA-binding;
CC Methionine biosynthesis; Repressor; Transcription;
CC Transcription regulation.
CC SQ SEQUENCE 105 AA; 12119 MW; 9443B57B03EB2DE4 CRC64;
Query Match 52.9%; Score 45; DB 1; Length 105;
Best Local Similarity 58.3%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 WNGESEKPYDDH 13
DB 4 WNGEYISPYAEH 15
[1] ||||| :|
[1] ||||| :|
RESULT 32
METJ VIBVU STANDARD; PRT; 105 AA.
ID METJ VIBVU
AC Q8DCN7;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Met repressor (Met regulon regulatory protein metJ).
GN Name=metJ; OrderedLocustNames=VJ11382;
OS Vibrio vulnificus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=672;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=14656965; DOI=10.1101/gr.1295503;
RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,
RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-F.;
RT "Comparative genome analysis of Vibrio vulnificus, a marine
RT pathogen.";
RL Genome Res. 13:2577-2587(2003).
CC -!- FUNCTION: This regulatory protein, when combined with SAM (S-
CC adenosylmethionine) represses the expression of the methionine
CC regulon and of enzymes involved in SAM synthesis (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- DOMAIN: Does not bind DNA by an helix-turn-helix motif.
CC -!- SIMILARITY: Belongs to the metJ family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC ENBL; BA000037; BAC95773.1; -; Genomic_DNA.
CC SMR; Q7MH63; 2-104.
CC HAMAP; MF 00744; -; 1.
CC InterPro; IPR002084; MetJ.
CC Pfam; PF01340; MetJ; 1.
CC PIRSF; PIRSF003191; MetJ; 1.
CC ProDom; PD020365; MetJ; 1.

```

```

CC -!- DOMAIN: Does not bind DNA by an helix-turn-helix motif.
CC -!- SIMILARITY: Belongs to the metJ family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC ENBL; AE016801; AAO09813.1; -; Genomic_DNA.
CC HSP; P08338; 1CMB.
CC SMR; Q8DCN7; 2-104.
CC HAMAP; MF 00744; -; 1.
CC InterPro; IPR002084; MetJ.
CC Pfam; PF01340; MetJ; 1.
CC ProDom; PD020365; MetJ; 1.
CC Amino-acid biosynthesis; Complete proteome; DNA-binding;
CC Methionine biosynthesis; Repressor; Transcription;
CC Transcription regulation.
CC SQ SEQUENCE 105 AA; 12063 MW; 28EF6669D924CA55 CRC64;
Query Match 52.9%; Score 45; DB 1; Length 105;
Best Local Similarity 58.3%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 WNGESEKPYDDH 13
DB 4 WNGEYISPYAEH 15
[1] ||||| :|
[1] ||||| :|
RESULT 33
METJ VIBVU STANDARD; PRT; 105 AA.
ID METJ VIBVU
AC Q7MH63;
DT 29-MAR-2004 (Rel. 43, Created)
DT 29-MAR-2004 (Rel. 43, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Met repressor (Met regulon regulatory protein metJ).
GN Name=metJ; OrderedLocustNames=VW3009;
OS Vibrio vulnificus (strain Y0016).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=196600;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RX PubMed=14656965; DOI=10.1101/gr.1295503;
RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,
RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,
RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-F.;
RT "Comparative genome analysis of Vibrio vulnificus, a marine
RT pathogen.";
RL Genome Res. 13:2577-2587(2003).
CC -!- FUNCTION: This regulatory protein, when combined with SAM (S-
CC adenosylmethionine) represses the expression of the methionine
CC regulon and of enzymes involved in SAM synthesis (By similarity).
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- DOMAIN: Does not bind DNA by an helix-turn-helix motif.
CC -!- SIMILARITY: Belongs to the metJ family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC ENBL; BA000037; BAC95773.1; -; Genomic_DNA.
CC SMR; Q7MH63; 2-104.
CC HAMAP; MF 00744; -; 1.
CC InterPro; IPR002084; MetJ.
CC Pfam; PF01340; MetJ; 1.
CC PIRSF; PIRSF003191; MetJ; 1.
CC ProDom; PD020365; MetJ; 1.

```

KW Amino-acid biosynthesis; Complete proteome; DNA-binding;  
 KW Methionine biosynthesis; Repressor; Transcription;  
 KW Transcription regulation;  
 SQ SEQUENCE 105 AA; 12063 MW; 28EF6669D924CA55 CRC64;

Query Match 52.9%; Score 45; DB 1; Length 105;  
 Best Local Similarity 58.3%; Pred. No. 21;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 WNGESEKPYDDH 13  
 ||||| ||:|  
 Db 4 WNGEYISPYAEH 15

## RESULT 34

IDENTITY METJ\_VIBPA STANDARD; PRT; 106 AA.  
 AC Q87L49;  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 13-SEP-2005 (Rel. 48, Last annotation update)  
 DE Met repressor (Met regulon regulatory protein metJ).  
 GN Name:metJ; OrderedLocusNames=VP2766;  
 OS Vibrio parahaemolyticus.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
 OC Vibrionaceae; Vibrio.  
 OX NCBI\_TaxID=670;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=RIMD 2210633 / Serotype O3:K6;  
 RX MEDLINE=22508454; PubMed=12620739; DOI=10.1016/S0140-6736(03)12659-1;  
 RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,  
 RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,  
 RA Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T.;  
 RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism  
 distinct from that of V. cholerae";  
 RL Lancet 361:743-749(2003).

CC -1- FUNCTION: This regulatory protein, when combined with SAM (S-adenosylmethionine) represses the expression of the methionine regulon and of enzymes involved in SAM synthesis (By similarity).  
 CC -1- SUBUNIT: Homodimer (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- DOMAIN: Does not bind DNA by an helix-turn-helix motif.  
 CC -1- SIMILARITY: Belongs to the metJ family.

-----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

CC EMBL; BA000031; BAC61029.1; -; Genomic\_DNA.  
 CC HSSP; P08338; ICMB.  
 CC SNR; Q87L49; 2-104.

DR HAMAP; MF 00744; -; 1.

DR InterPro; IPR002084; MetJ.

DR Pfam; PF01340; MetJ; 1.

DR PIRSF; PIRSF003191; MetJ; 1.

DR ProDom; PD020365; MetJ; 1.

KW Amino-acid biosynthesis; Complete proteome; DNA-binding;  
 KW Methionine biosynthesis; Repressor; Transcription;  
 KW Transcription regulation;  
 SQ SEQUENCE 106 AA; 12132 MW; 06E95F74F9D924CF CRC64;

Query Match 52.9%; Score 45; DB 1; Length 106;  
 Best Local Similarity 58.3%; Pred. No. 22;  
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 2 WNGESEKPYDDH 13  
 ||||| ||:|  
 Db 4 WNGEYISPYAEH 15

## RESULT 35

IDENTITY Q7KWW8\_DICDI PRELIMINARY; PRT; 373 AA.  
 AC Q7KWW8;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)  
 DE D1-N-acetylcholinesterase (EC 3.2.1.-) (Hypothetical protein).  
 GN ORFNames=DD80168307;  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=AX4;  
 RX MEDLINE=22092622; PubMed=12097910; DOI=10.1038/nature00847;  
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J.A.,  
 RA Bankier A.T., Dear P.H., Lehmann R., Baumgart C., Parra G.,  
 RA Abril J.F., Guigo R., Kumpf K., Tunggal B., Cox E., Quail M.A.,  
 RA Platzer M., Rosenthal A., Noegel A.A.;  
 RT "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";  
 RL Nature 418:79-85(2002).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=AX4;  
 RC Baumgart C.;  
 RA Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RP NUCLEOTIDE SEQUENCE.

CC Eichinger L., Pachebat J.A., Gloeckner G., Rajandream M.-A.,  
 CC Sugcang R., Berriman M., Song J., Olsen R., Szafranski K., Xu Q.,  
 CC Tunggal B., Kummerfeld S., Madera M., Konfortov B.A., Rivero F.,  
 CC Banker A.T., Lehmann R., Hamlin N., Davies R., Gaudet P., Fey P.,  
 CC Pilcher K., Chen G., Saunders D., Sodeguren E., Davis P.,  
 CC Kerhrou A., Nie X., Hall N., Anjard C., Hemphill L., Bason N.,  
 CC Farbrother P., Desany B., Just E., Morio T., Rost R., Churcher C.,  
 CC Cooper J., Haydock S., van Driessche N., Cronin A., Goodhead I.,  
 CC Muzny D., Mourier T., Pain A., Lu M., Harper D., Lindsay R.,  
 CC Hauser H., James K., Quiles M., Mohan M.B., Saito T., Buchrieser C.,  
 CC Wardrop A., Felder M., Thangavelu M., Johnson D., Knights A.,  
 CC Loulès H., Mungall K., Oliver K., Price C., Quail M.A., Sanders M.,  
 CC Urushihara H., Hernandez J., Rabinowitsch E., Steffen D., Tivey A.,  
 CC Ma J., Kohara Y., Sharp S., Simmonds M., Spiegler S., Tivey A.,  
 CC Sugano S., White B., Walker D., Woodward J., Winckler T., Tanaka Y.,  
 CC Shaulsky G., Schleicher M., Weinstein G., Rosenthal A., Cox E.C.,  
 CC Chisholm R.L., Gibbs R., Loomis W.F., Platzer M., Kay R.R.,  
 CC Williams J., Dear P.H., Noegel A.A., Barrell B., Kuspa A.;  
 RT "The genome of the social amoeba Dictyostelium discoideum.";  
 RL Nature 0:0-0(2005).

DR EMBL; AC115581; AAS38675.1; -; Genomic\_DNA.

DR EMBL; AAFI01000042; EAL68899.1; -; Genomic\_DNA.

DR GO; GO:0016787; F:hydrolase activity; IEA.

DR GO; GO:0008152; P:metabolism; IEA.

DR InterPro; IPR011583; Chitinase-II.

DR InterPro; IPR001223; Glyco\_hydro\_18.

DR Pfam; PF00704; Glyco\_hydro\_18; 1.

DR SMART; SM00636; Glyco\_18; 1.

KW Hypothetical protein.

SQ SEQUENCE 373 AA; 42433 MW; B55257242C3C6863 CRC64;

Query Match 52.9%; Score 45; DB 2; Length 373;  
 Best Local Similarity 70.0%; Pred. No. 89;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 QWNGESEKPY 10

Db 298 QWDSSESPY 307

## RESULT 36

IDENTITY SAHH\_SULTO STANDARD; PRT; 415 AA.

```

AC Q975T0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Adenosylhomocysteinase (EC 3.3.1.1) (S-adenosyl-L-homocysteine
DE hydrolase) (AdoHcyase).
GN Name=ahcy; OrderedLocusNames=ST0342;
OS Sulfolobus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=JCM 10545 / 7;
RX MEDLINE=21456156; PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-homocysteine + H(2)O = L-
CC homocysteine + adenosine.
CC -1- COFACTOR: Binds 1 NAD per subunit (By similarity).
CC -1- PATHWAY: Activated methyl cycle.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the adenosylhomocysteinase family.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; BA000023; BAB65319.1; -; Genomic_DNA.
DR HAMAP; P23526; LU14.
DR InterPro; IPR000043; Ad_hcy_hydrolase.
DR PANTHER; PTHR11784; Ad_hcy_hydrolase; 1.
DR Pfam; PF05221; AdoHcyase; 1.
DR Pfam; PF00670; AdoHcyase_NAD; 1.
DR TIGRfams; TIGR00936; ahcy; 1.
DR PROSITE; PS00738; ADOHCYASE_1; 1.
DR PROSITE; PS00739; ADOHCYASE_2; 1.
KW Complete proteome; Hydrolase; NAD; One-carbon metabolism.
FT REGION 174 341 NAD binding (By similarity).
FT BINDING 53 53 Substrate (By similarity).
FT BINDING 124 124 Substrate (By similarity).
FT BINDING 147 147 Substrate (By similarity).
FT BINDING 177 177 Substrate (By similarity).
FT BINDING 181 181 Substrate (By similarity).
SQ SEQUENCE 415 AA; 46207 MW; EDP3DF7F369BA5D CRC64;

Query Match 52.9%; Score 45; DB 1; Length 415;
Best Local Similarity 53.8%; Pred. No. 1e+02;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 WNGESEKPYDDHL 14
DB 98 WRGETEKDYDNI 110
| ||| ||| |:::

RESULT 37
ID Q4JAZ7 SULAC PRELIMINARY; PRT; 415 AA.
AC Q4JAZ7;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Adenosylhomocysteinase (EC 3.3.1.1).

```

```

GN Name=ahcy; OrderedLocusNames=Saci_0646;
OS Sulfolobus acidocaldarius.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2285;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;
RX PubMed=15995215; DOI=10.1128/JB.187.14.4992-4999.2005;
RA Chen L., Bruegger K., Skovgaard M., Redder P., She Q., Torarinsson E.,
RA Greve B., Awayez M., Zibat A., Klenk H.-P., Garrett R.A.;
RT "The genome of Sulfolobus acidocaldarius, a model organism of the
RT Crenarchaeota.";
RL J. Bacteriol. 187:4992-4999(2005).
DR EMBL; CP000077; AAY80032.1; -; Genomic_DNA.
KW Complete proteome; Hydrolase.
SQ SEQUENCE 415 AA; 46176 MW; 298B46FE0235CC9A CRC64;

Query Match 52.9%; Score 45; DB 2; Length 415;
Best Local Similarity 53.8%; Pred. No. 1e+02;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 2 WNGESEKPYDDHL 14
DB 98 WRGETEKDYDNI 110
| ||| ||| |:::

RESULT 38
ID Q6NXJ0 MOUSE PRELIMINARY; PRT; 1187 AA.
AC Q6NXJ0;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein D8Ert594e.
GN Name=D8Ert594e;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi;
OC Muroidae; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnurch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=C57BL/6; TISSUE=Brain;
RG NIH MGC Project;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC067050; AAH67050.1; -; mRNA.
DR Ensembl; ENSMUSG00000031563; Mus musculus.
DR MGI; MGI:1261872; D8Ert594e.

```

```

DR InterPro; IPR000008; C2.
DR InterPro; IPR001202; WW_Rep5_WWP.
DR Pfam; PF00168; C2; 1.
DR Pfam; PF00397; WW; 2.
DR SMART; SM00456; WW; 2.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS50020; WW_DOMAIN_2; 2.
KW Hypothetical protein.
SQ SEQUENCE 1187 AA; 132620 MW; 40FB67C9D1D3B296 CRC64;

Query Match          52.9%; Score 45; DB 2; Length 1187;
Best Local Similarity 57.1%; Pred. No. 3.3e+02;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 QWNGESEKPYDDHL 14
   ||||| :|||
Db 91 QWRGEQEKMLKDYL 104

RESULT 39
ID O71964_9GEMI PRELIMINARY; PRT; 85 AA.
AC O71964;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AC4.
GN Name=AC4;
OS Sida golden mosaic virus.
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=51034;
RN [1]
RP Abouzid A.M., Polston J.E., Hiebert E.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF049336; AAC05155.1; -; Genomic_DNA.
DR InterPro; IPR002488; Gemini_C4.
DR Pfam; PF01492; Gemini_C4; 1.
SQ SEQUENCE 85 AA; 9458 MW; E6733C13AF280AD6 CRC64;

Query Match          51.8%; Score 44; DB 2; Length 85;
Best Local Similarity 57.1%; Pred. No. 25;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 QWNGESEKPYDDHL 14
   ||||| :|||
Db 58 QWNGESSRSTADQL 71

RESULT 40
Q5QV41 IDILO PRELIMINARY; PRT; 107 AA.
AC Q5QV41;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Transcriptional regulator of met regulon (Repressor).
GN Name=metJ; OrderedLocusNames=IL2465;
OS Idiomarina loihiensis
OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
OC Idiomarinaceae; Idiomarina.
OX NCBI_TaxID=135577;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=L2-TR / DSM 15497 / ATCC BAA-735;
RX PubMed=15596722; DOI=10.1073/pnas.0407638102;
RA Hou S., Saw J.H., Lee K.S., Freitas T.A., Belisle C., Kawarabayasi Y.,
RA Donachie S.P., Pikina A., Galperin M.Y., Koonin E.V., Makarova K.S.,
RA Onelchenko M.V., Sorokin A., Wolf Y.I., Li Q.X., Keum Y.S.,
RA Campbell S., Denery J., Aizawa S.-I., Shibata S., Malahoff A.,
RA Alam M.;
RT "genome sequence of the deep-sea gamma-proteobacterium Idiomarina
loihiensis reveals amino acid fermentation as a source of carbon and

```

```

RT energy.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:18036-18041(2004).
DR EMBL; AE017340; AAV83297.1; -; Genomic_DNA.
DR SMR; Q5QV41; 2-97.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006555; P:methionine metabolism; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR002084; MetJ.
DR Pfam; PF01340; MetJ; 1.
DR ProDom; PD020365; MetJ; 1.
KW Complete proteome.
SQ SEQUENCE 107 AA; 12367 MW; 11664CC9658C6CC4 CRC64;

Query Match          51.8%; Score 44; DB 2; Length 107;
Best Local Similarity 53.8%; Pred. No. 32;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QWNGESEKPYDDH 13
   :||| :|||
Db 2 KNGEYIYPAEH 14

RESULT 41
Q9ZBK4 STRCO PRELIMINARY; PRT; 170 AA.
ID Q9ZBK4_STRCO PRELIMINARY;
AC Q9ZBK4;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein SCO6470
GN OrderedLocusNames=SCO6470; ORFNames=SC9C7.06c;
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Kieser H.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Lake L., Murphy L.D., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939127; CAA22718.1; -; Genomic_DNA.
DR PIR; T35957; T35957.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR002539; MacC dehydratase.
DR Pfam; PF01575; MacC dehydratase; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 170 AA; 19254 MW; 69EBC3AA226FF3A9 CRC64;

Query Match          51.8%; Score 44; DB 2; Length 170;
Best Local Similarity 53.8%; Pred. No. 54;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 WNGESEKPYDDHL 14
   | :| :| :| :|
Db 19 WPGKTVTEYDDHL 31

RESULT 42
Q82LU6 STRAW PRELIMINARY; PRT; 175 AA.
ID Q82LU6_STRAW PRELIMINARY;
AC Q82LU6;

```

DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GS OrderedLocusNames=SAV1914;  
 OS Streptomyces avermitilis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=33903;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
 RX MEDLINE=22608306; PubMed=12692562; DOI=10.1038/nbt820;  
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
 RA Sakaki Y., Hattori M., Omura S.;  
 RT "Complete genome sequence and comparative analysis of the industrial  
 RT microorganism Streptomyces avermitilis";  
 RL Nat. Biotechnol. 21:526-531(2003).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
 RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;  
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
 RT "Genome sequence of an industrial microorganism Streptomyces  
 RT avermitilis: deducing the ability of producing secondary  
 RT metabolites";  
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
 DR EMBL; BA000030; BAC69625.1; -; Genomic DNA.  
 DR GO; GO:0016491; P:oxidoreductase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR002539; MAO\_cdehydratas.  
 DR Pfam; PF01575; MAO\_cdehydratas; 1.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 175 AA; 19750 MW; 07E6B371CCAD378A CRC64;

Query Match 51.8%; Score 44; DB 2; Length 175;  
 Best Local Similarity 53.8%; Pred. No. 56;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14  
 |:::|||||  
 Db 19 WPKCTVTEYDDHL 31

RESULT 43  
 Q6BXN6\_DEBHA  
 ID Q6BXN6\_DEBHA PRELIMINARY; PRT; 292 AA.  
 AC Q6BXN6;  
 DT 25-OCT-2004 (TREMBlrel. 28, Created)  
 DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 DE Debaryomyces hansenii chromosome B of strain CBS767 of Debaryomyces  
 DE hansenii.  
 GN OrderedLocusNames=DEHA0B01540g;  
 OS Debaryomyces hansenii (Yeast) [Torulaspora hansenii].  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
 OX NCBI\_TaxID=4959;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=ATCC 36239 / CBS 767;  
 RX PubMed=15229592; DOI=10.1038/nature02579;  
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neugeglise C., Talla E.,  
 RA Geoffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
 RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,  
 RA Boisrame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,  
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
 RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,

RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,  
 RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,  
 RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,  
 RA Wincker P., Souciet J.-L.;  
 RT "Genome evolution in yeasts";  
 RL Nature 430:35-44(2004).  
 DR EMBL; CR382134; CAG85019.1; -; Genomic\_DNA.  
 KW Complete proteome.  
 SQ SEQUENCE 292 AA; 33410 MW; 4E67710DBEBFD63B CRC64;

Query Match 51.8%; Score 44; DB 2; Length 292;  
 Best Local Similarity 56.2%; Pred. No. 99;  
 Matches 9; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 1 QWNGES--EKPYDDHL 14  
 ||||:|:|||||  
 Db 126 QWNENNDEKKWDDEL 141

RESULT 44  
 Q4RLR4\_TETNG  
 ID Q4RLR4\_TETNG PRELIMINARY; PRT; 378 AA.  
 AC Q4RLR4;  
 DT 13-SEP-2005 (TREMBlrel. 31, Created)  
 DT 13-SEP-2005 (TREMBlrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TREMBlrel. 31, Last annotation update)  
 DE Chromosome 10 SCAF15019, whole genome shotgun sequence.  
 GN ORFNAMES=GSTENG00032376001;  
 OS Tetraodon nigroviridis (Green puffer).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Tetraodontidae; Tetraodon.  
 OX NCBI\_TaxID=99883;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA Jaillon O., Aury J.-M., Brunet F., Petit J.-L., Stange-Thomann N.,  
 RA Mauceli E., Bouneau L., Fischer C., Ozouf-Costaz C., Bernot A.,  
 RA Nicaud S., Jaffe D., Fisher S., Lutfalla G., Dossat C., Segurens B.,  
 RA Dasilva C., Salanoubat M., Levy M., Boudet N., Castellano S.,  
 RA Anthouard V., Jubin C., Castelli V., Katinka M., Vacherie B.,  
 RA Biemont C., Skalli Z., Cattolico L., Poulain J., De Berardinis V.,  
 RA Cruaud C., Duprat S., Brottier P., Coutanceau J.-P., Gouzy J.,  
 RA Parra G., Lardier G., Chappie C., McKernan K.J., McEwan P., Bosak S.,  
 RA Kellis M., Volff J.N., Guigo R., Zody M.C., Mesirov J.,  
 RA Lindblad-Toh K., Birren B., Nusbaum C., Kahn D., Robinson-Rechavi M.,  
 RA Laudet V., Schachter V., Quetier F., Saurin W., Scarpelli C.,  
 RA Wincker P., Lander E.S., Weissenbach J., Roest Crolius H.;  
 RT "Genome duplication in the teleost fish Tetraodon nigroviridis reveals  
 RT the early vertebrate proto-karyotype";  
 RL Nature 431:946-957(2004).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RG Genoscope; Whitehead Institute Centre for Genome Research;  
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.  
 CC -i- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; CAAB01015019; CAG10668.1; -; Genomic DNA.  
 SQ SEQUENCE 378 AA; 43844 MW; E58C371D58BC7BA8 CRC64;

Query Match 51.8%; Score 44; DB 2; Length 378;  
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGSEKPY 10  
 ||||:|:|||||  
 Db 289 WNGSTEQPY 297

RESULT 45  
 Q874K7\_9BASI



ID Q874K7\_9BASI PRELIMINARY; PRT; 411 AA.  
 AC Q874K7;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Epoxide hydrolase.  
 GN Names=EPHL;  
 OS Rhodospiridium paludigenum.  
 OC Eukaryota; Fungi; Basidiomycota; Urediniomycetes;  
 OC Microbotryomycetidae; Sporidiobolales; Rhodospiridium.  
 OX NCBI\_TaxID=96838;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=CSS565;  
 RA Labuschagne M., Albertyn J., Botes A.L.;  
 RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: An epoxide + H(2)O = a glycol.  
 CC -1- SUBCELLULAR LOCATION: Membrane-bound on microsomes (By similarity).  
 CC similarity).  
 DR EMBL; AY230137; AA072994.1; -; Genomic\_DNA.  
 DR HSSP; Q9UR30; 1007.  
 DR GO; GO:0005783; C:cytoplasmic reticulum; IEA.  
 DR GO; GO:0016021; C:integral to membrane; IEA.  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:0005792; C:microsome; IEA.  
 DR GO; GO:0004301; F:epoxide hydrolase activity; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0019439; P:aromatic compound catabolism; IEA.  
 DR GO; GO:0009636; P:response to toxin; IEA.  
 DR InterPro; IPR000073; A/b hydrolase.  
 DR InterPro; IPR010497; EH\_N.  
 DR InterPro; IPR000639; Epox hydrolase.  
 DR InterPro; IPR006025; Pept M\_Zn\_BS.  
 DR InterPro; IPR000379; Ser esters.  
 DR Pfam; PF00561; Abhydrolase\_1; 1.  
 DR Pfam; PF06441; EHN\_1.  
 DR PRINTS; PR00412; EPOXYDRLASE.  
 DR PROSITE; PS00142; ZINC PROTEASE; UNKNOWN 1.  
 KW Aromatic hydrocarbons catabolism; Endoplasmic reticulum; Hydrolase;  
 KW Microsome; Transmembrane.  
 SQ SEQUENCE 411 AA; 46172 MW; 8D43FC033B269B2B CRC64;  
  
 Query Match 51.8%; Score 44; DB 2; Length 411;  
 Best Local Similarity 42.9%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
  
 QY 1 QWNGESEKPYDDHL 14  
 :|:|:|:|  
 Db 68 EWRGQDQKQLQDHL 81  
  
 RESULT 46  
 O54549\_HALSA PRELIMINARY; PRT; 416 AA.  
 ID O54549;  
 AC O54549;  
 DT 01-JUN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocNames=VNG5070, VNG5226;  
 OS Halobacterium salinarum (Halobacterium halobium).  
 OG Plasmid pNRC100.  
 OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
 OC Halobacteriaceae; Halobacterium.  
 OX NCBI\_TaxID=2242;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RC STRAIN=NRC-1 / ATCC 700922 / JCM 11081; PLASMID=pNRC100;  
 RX MEDLINE=20504483; PubMed=11016950; DOI=10.1073/pnas.190337797;  
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Bergquist B., Pan M.,  
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,  
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,  
 RA Leithauser B., Kellier K., Cruz R., Danson M.J., Hough D.W.,

RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,  
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,  
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,  
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;  
 RT "Genome sequence of Halobacterium species NRC-1";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=NRC-1 / ATCC 700922 / JCM 11081; PLASMID=pNRC100;  
 RX MEDLINE=99063795; PubMed=9847077;  
 RA Ng W.V., Ciuffo S.A., Smith T.M., Bumgarner R.E., Baskin D., Faust J.,  
 RA Hall B., Loretz C., Seto J., Slagel J., Hood L., DasSarma S.;  
 RT "Snapshot of a large dynamic replicon in a halophilic archaeon:  
 RT megaplasmid or minichromosome?";  
 RL Genome Res. 8:1131-1141(1998).  
 DR EMBL; AF016485; AAC82943.1; -; Genomic DNA.  
 DR EMBL; AF016485; AAC82943.1; -; Genomic DNA.  
 DR PIR; T08267; T08267.  
 KW Complete proteome; Hypothetical protein; Plasmid.  
 SQ SEQUENCE 416 AA; 47068 MW; 7CE5AAC3A9C9C61 CRC64;  
  
 Query Match 51.8%; Score 44; DB 2; Length 416;  
 Best Local Similarity 53.8%; Pred. No. 1.5e+02;  
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 1 QWNGESEKPYDDH 13  
 :|:|:|:|  
 Db 214 QEDGDEERQYDDH 226  
  
 RESULT 47  
 Q55JV8\_CRYNE PRELIMINARY; PRT; 453 AA.  
 ID Q55JV8;  
 AC Q55JV8;  
 DT 13-SEP-2005 (TrEMBLrel. 31, Created)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)  
 DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=CNBK2250;  
 OS Cryptococcus neoformans var. neoformans B-3501A.  
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;  
 OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.  
 OX NCBI\_TaxID=283643;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=B-3501A;  
 RA Fung E., Hyman R.W., Rowley D., Bruno D., Miranda M., Fukushima M.,  
 RA Wickes B.L., Fu J., Davis R.W.;  
 RT "Cryptococcus neoformans serotype D sequencing";  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AA01000052; EAL18207.1; -; Genomic\_DNA.  
 KW Hypothetical protein.  
 SQ SEQUENCE 453 AA; 51310 MW; 022BB2D450365088 CRC64;  
  
 Query Match 51.8%; Score 44; DB 2; Length 453;  
 Best Local Similarity 61.5%; Pred. No. 1.6e+02;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
  
 QY 2 WNGESEKPYDDHL 14  
 :|:|:|:|  
 Db 424 WHGESEEMDDVL 436  
  
 RESULT 48  
 Q5K9P2\_CRYNE PRELIMINARY; PRT; 453 AA.  
 ID Q5K9P2;  
 AC Q5K9P2;  
 DT 10-MAY-2005 (TrEMBLrel. 30, Created)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)  
 DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)

```

DE Expressed protein.
GN ORFNAMES=CNK01280;
OS Cryptococcus neoformans var. neoformans JEC21.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
OX NCBI_TaxID=214684;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=JEC21;
RA Loftus B., Amedeo P., Roncaglia P., Vamathevan J., Utterback T.,
RA Van Aken S., Fraser C.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=JEC21; DOI=10.1126/science.1103773;
RX PubMed=15653466;
RA Loftus B.J., Fung E., Roncaglia P., Rowley D., Amedeo P., Bruno D.,
RA Vamathevan J., Miranda M., Anderson I.J., Fraser J.A., Allen J.E.,
RA Bosdet I.E., Brent M.R., Chiu R., Doering T.L., Donlin M.J.,
RA D'Souza C.A., Fox D.S., Grinberg V., Fu J., Fukushima M., Haas B.J.,
RA Huang J.C., Janbon G., Jones S.J.M., Koo H.L., Krzywinski M.I.,
RA Kwon-Chung K.J., Lengler K.B., Maiti R., Marra M.A., Marra R.E.,
RA Mathewson C.A., Mitchell T.G., Perlea M., Riggs P.R., Salzberg S.L.,
RA Schein J.E., Shvartsbeyn A., Shin H., Shumway M., Specht C.A.,
RA Suh B.B., Tenney A., Utterback T.R., Wickes B.L., Wortman J.R.,
RA Wye N.H., Kronstad J.W., Lodge J.K., Heitman J., Davis R.W.,
RA Fraser C.M., Hyman R.;
RT "The genome of the basidiomycetous yeast and human pathogen
RT Cryptococcus neoformans.";
RL Science 307:1321-1324 (2005).
DR EMBL; AE017351; AAW46218.1; -; Genomic_DNA.
KW Complete proteome.
SQ SEQUENCE 453 AA; 51310 MW; 022EB2D450365088 CRC64;

Query Match 51.8%; Score 44; DB 2; Length 453;
Best Local Similarity 61.5%; Pred. No. 1.6e+02;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 WNGESEKPYDDHL 14
|:||||: |||
DB 424 WHGESEEMDDVL 436

RESULT 49
O7UKJ8 RHOB
ID O7UKJ8_RHOB PRELIMINARY; PRT; 523 AA.
AC O7UKJ8;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Bifunctional purine biosynthesis protein purH (EC 2.1.2.3)
DE (EC 3.5.4.10).
GN Name=purH; OrderedLocusNames=RB10113;
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416; DOI=10.1073/pnas.1431443100;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL; BX294150; CAD76635.1; -; Genomic_DNA.
DR HSSP; P31939; 1PKX.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0003937; F:IMP cyclonhydrolase activity; IEA.
DR GO; GO:0004643; F:phosphoribosylaminimidazolecarboxamide for. . .; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.

```

```

DR GO; GO:0006164; P:purine nucleotide biosynthesis; IEA.
DR InterPro; IPR002695; AICARFT_IMPCHas.
DR DR
DR Pfam; PF01808; AICARFT_IMPCHas; 1.
DR Pfam; PF02142; MGS; 1.
DR ProDom; PD004666; AICARFT_IMPCHas; 1.
DR TIGRFAMs; TIGR00355; purH; 1.
KW Complete proteome; Hydrolase; Transferase.
SQ SEQUENCE 523 AA; 56202 MW; 28DEDD894009E924 CRC64;

Query Match 51.8%; Score 44; DB 2; Length 523;
Best Local Similarity 57.1%; Pred. No. 1.9e+02;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 OWNGESEKPYDDHL 14
|:||||: |||
DB 397 QWNTVETPVDL 410

RESULT 50
O6BWQ7 DEBHA
ID O6BWQ7_DEBHA PRELIMINARY; PRT; 669 AA.
AC O6BWQ7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to CA3316|IPF14542 Candida albicans IPF14542 unknown
DE function.
GN OrderedLocusNames=DEHA0809680g;
OS Debaryomyces hansenii (Yeast) (Torulasporea hansenii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OX NCBI_TaxID=4959;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 36239 / CBS 767;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anhouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boiserame A., Boyer J., Cattolico L., Confantolero F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Fellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44 (2004).
DR EMBL; CR382134; CAG85366.1; -; Genomic_DNA.
KW Complete proteome.
SQ SEQUENCE 669 AA; 76112 MW; 37590D8EA28B684B CRC64;

Query Match 51.8%; Score 44; DB 2; Length 669;
Best Local Similarity 50.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 WNGESEKPYDDH 13
|:||||: |||
DB 642 WNGLPQEPHDDN 653

Search completed: February 10, 2006, 23:34:44
Job time : 316 secs

```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:35:03 ; Search time 50 Seconds  
(without alignments)  
23.149 Million cell updates/sec

Title: US-10-797-821-28

Perfect score: 85

Sequence: 1 QWNGSEKPYDDHL 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep:\*

2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep:\*

3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep:\*

4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*

5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep:\*

6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85	100.0	14	1	US-08-057-162B-5
2	85	100.0	15	2	US-09-230-049A-4
3	85	100.0	15	2	US-09-230-049A-8
4	85	100.0	15	2	US-09-230-049A-9
5	67	78.8	15	2	US-09-230-049A-6
6	67	78.8	1375	2	US-09-210-361-4
7	67	78.8	1375	2	US-09-740-274-4
8	67	78.8	1475	2	US-09-007-599-2
9	67	78.8	1475	2	US-09-210-361-2
10	67	78.8	1475	2	US-09-740-274-2
11	64	75.3	15	2	US-09-230-049A-5
12	47.5	55.9	357	2	US-09-543-681A-4957
13	47	55.3	106	2	US-09-543-681A-8307
14	46	54.1	504	2	US-09-134-001C-2980
15	42	49.4	187	2	US-09-489-031A-12202
16	42	49.4	271	2	US-09-252-991A-30733
17	42	49.4	424	2	US-09-583-110-3119
18	42	49.4	426	2	US-09-107-433-3690
19	41	48.2	473	2	US-09-543-681A-4480
20	40	47.1	189	2	US-09-270-767-34087
21	40	47.1	189	2	US-09-270-767-49304
22	40	47.1	252	2	US-09-134-000C-3729
23	40	47.1	319	2	US-09-710-279-2760
24	40	47.1	443	2	US-09-134-001C-3183
25	40	47.1	532	2	US-09-487-558B-360
26	39	45.9	107	2	US-09-270-767-33478
27	39	45.9	309	2	US-09-710-279-356
					Sequence 5, Appli
					Sequence 4, Appli
					Sequence 8, Appli
					Sequence 9, Appli
					Sequence 6, Appli
					Sequence 4, Appli
					Sequence 4, Appli
					Sequence 8307, Ap
					Sequence 2980, Ap
					Sequence 12202, A
					Sequence 30733, A
					Sequence 3119, Ap
					Sequence 3690, Ap
					Sequence 4480, Ap
					Sequence 34087, A
					Sequence 49304, A
					Sequence 3729, Ap
					Sequence 2760, Ap
					Sequence 3183, Ap
					Sequence 360, App
					Sequence 33478, A
					Sequence 356, App

28	39	45.9	313	2	US-09-134-001C-4371	Sequence 4371, Ap
29	39	45.9	594	2	US-09-117-415B-2	Sequence 2, Appli
30	39	45.9	599	2	US-09-248-796A-17744	Sequence 17744, A
31	39	45.9	613	2	US-09-117-415B-22	Sequence 22, Appl
32	39	45.9	631	2	US-09-117-415B-18	Sequence 18, Appl
33	39	45.9	631	2	US-09-117-415B-20	Sequence 20, Appl
34	39	45.9	649	2	US-09-117-415B-16	Sequence 16, Appl
35	39	45.9	1602	2	US-09-269-874A-7	Sequence 7, Appli
36	39	45.9	1621	2	US-09-269-874A-5	Sequence 5, Appli
37	39	45.9	1639	2	US-09-269-874A-3	Sequence 3, Appli
38	38	44.7	124	2	US-09-134-000C-4510	Sequence 4510, Ap
39	38	44.7	144	2	US-09-540-236-2975	Sequence 2975, Ap
40	38	44.7	269	2	US-09-543-681A-7478	Sequence 7478, Ap
41	38	44.7	388	2	US-09-248-796A-18350	Sequence 18350, A
42	38	44.7	443	2	US-09-252-991A-21256	Sequence 21256, A
43	38	44.7	480	2	US-09-248-796A-17167	Sequence 17167, A
44	38	44.7	532	2	US-09-389-956-80	Sequence 80, Appl
45	38	44.7	638	2	US-09-248-796A-18466	Sequence 18466, A
46	38	44.7	1095	2	US-08-851-567B-34	Sequence 34, Appl
47	37.5	44.1	349	2	US-09-270-767-43051	Sequence 43051, A
48	37.5	44.1	412	2	US-09-445-472-1	Sequence 1, Appli
49	37.5	44.1	412	2	US-10-090-624-1	Sequence 1, Appli
50	37.5	44.1	522	2	US-08-894-818B-3	Sequence 3, Appli
51	37.5	44.1	522	2	US-09-445-472-4	Sequence 4, Appli
52	37.5	44.1	522	2	US-10-090-624-4	Sequence 4, Appli
53	37.5	44.1	522	2	US-09-841-553-3	Sequence 3, Appli
54	37.5	44.1	654	2	US-08-894-818B-35	Sequence 35, Appl
55	37.5	44.1	654	2	US-09-445-472-16	Sequence 16, Appl
56	37.5	44.1	654	2	US-10-090-624-16	Sequence 16, Appl
57	37.5	44.1	654	2	US-09-841-553-35	Sequence 35, Appl
58	37.5	44.1	659	2	US-08-894-818B-5	Sequence 5, Appli
59	37.5	44.1	659	2	US-09-841-553-3	Sequence 5, Appli
60	37	43.5	87	2	US-09-621-976-5345	Sequence 5345, Ap
61	37	43.5	99	2	US-08-930-894-8	Sequence 8, Appli
62	37	43.5	101	1	US-08-341-843B-14	Sequence 14, Appl
63	37	43.5	101	1	US-08-427-497B-19	Sequence 19, Appl
64	37	43.5	118	2	US-09-248-796A-28195	Sequence 28195, A
65	37	43.5	209	2	US-08-506-296B-61	Sequence 61, Appl
66	37	43.5	220	2	US-09-252-991A-21649	Sequence 21649, A
67	37	43.5	301	2	US-09-902-540-14155	Sequence 14155, A
68	37	43.5	302	2	US-09-457-046B-8	Sequence 8, Appli
69	37	43.5	302	2	US-09-866-570B-8	Sequence 8, Appli
70	37	43.5	309	2	US-08-506-296B-60	Sequence 60, Appl
71	37	43.5	397	2	US-08-506-296B-59	Sequence 59, Appl
72	37	43.5	441	2	US-09-543-681A-5018	Sequence 5018, Ap
73	37	43.5	447	2	US-08-506-296B-73	Sequence 73, Appl
74	37	43.5	502	2	US-09-252-991A-23501	Sequence 23501, A
75	37	43.5	547	2	US-08-506-296B-72	Sequence 72, Appl
76	37	43.5	555	1	US-08-793-229-34	Sequence 34, Appl
77	37	43.5	555	2	US-09-285-957-34	Sequence 34, Appl
78	37	43.5	565	2	US-09-252-991A-27022	Sequence 27022, A
79	37	43.5	635	2	US-08-506-296B-71	Sequence 71, Appl
80	37	43.5	781	2	US-09-252-991A-23353	Sequence 23353, A
81	37	43.5	844	1	US-07-731-157A-6	Sequence 6, Appli
82	37	43.5	844	1	US-08-541-780-6	Sequence 6, Appli
83	37	43.5	911	2	US-09-074-579-4	Sequence 4, Appli
84	37	43.5	911	2	US-09-388-774-4	Sequence 4, Appli
85	37	43.5	947	2	US-09-252-991A-21335	Sequence 21335, A
86	37	43.5	1034	2	US-09-252-991A-20969	Sequence 20969, A
87	37	43.5	1041	2	US-09-252-991A-23266	Sequence 23266, A
88	37	43.5	1065	2	US-08-630-172-9	Sequence 9, Appli
89	37	43.5	1065	2	US-09-375-419-9	Sequence 21, Appl
90	37	43.5	1260	2	US-08-506-296B-21	Sequence 25089, A
91	36	42.4	92	2	US-09-248-796A-25089	Sequence 59170, A
92	36	42.4	117	2	US-09-270-767-59170	Sequence 59170, A
93	36	42.4	120	2	US-09-270-767-41469	Sequence 41469, A
94	36	42.4	147	2	US-09-270-767-33757	Sequence 33757, A
95	36	42.4	147	2	US-09-270-767-48974	Sequence 48974, A
96	36	42.4	163	2	US-08-858-207A-338	Sequence 338, App
97	36	42.4	243	2	US-09-270-767-41921	Sequence 41921, A
98	36	42.4	269	2	US-09-809-920-11	Sequence 11, Appl
99	36	42.4	283	2	US-09-952-689-5	Sequence 5, Appli
100	36	42.4	286	2	US-09-270-767-43773	Sequence 43773, A

101	36	42.4	292	2	US-09-372-422A-10	Sequence 10, Appl	174	35	41.2	514	2	US-09-134-000C-5666	Sequence 5666, Ap
102	36	42.4	316	2	US-09-248-796A-20451	Sequence 20451, A	175	35	41.2	515	2	US-09-248-796A-18264	Sequence 18264, A
103	36	42.4	344	2	US-09-583-110-3959	Sequence 3959, Ap	176	35	41.2	519	2	US-08-956-171B-5230	Sequence 5230, Ap
104	36	42.4	354	2	US-09-198-452A-504	Sequence 504, App	177	35	41.2	519	2	US-08-781-986A-5230	Sequence 5230, Ap
105	36	42.4	354	2	US-09-438-185A-471	Sequence 471, App	178	35	41.2	528	2	US-09-252-991A-32010	Sequence 32010, A
106	36	42.4	359	2	US-09-107-433-2646	Sequence 2646, App	179	35	41.2	550	2	US-09-199-637A-202	Sequence 202, App
107	36	42.4	414	2	US-09-248-796A-16117	Sequence 16117, A	180	35	41.2	558	1	US-08-320-559-30	Sequence 30, Appl
108	36	42.4	428	2	US-09-792-420-8	Sequence 8, Appl	181	35	41.2	568	2	US-08-545-860D-30	Sequence 30, Appl
109	36	42.4	440	2	US-09-543-681A-6033	Sequence 6033, Ap	182	35	41.2	568	4	PCT-US94-04496-30	Sequence 1114, Ap
110	36	42.4	448	2	US-09-248-796A-20279	Sequence 20279, A	183	35	41.2	568	4	PCT-US94-04496-30	Sequence 10, Appl
111	36	42.4	510	2	US-09-248-796A-20798	Sequence 20798, A	184	35	41.2	608	2	US-09-199-637A-201	Sequence 201, App
112	36	42.4	607	2	US-08-482-934A-2	Sequence 2, Appl	185	35	41.2	639	2	US-09-199-637A-200	Sequence 200, App
113	36	42.4	607	2	US-08-182-045-2	Sequence 2, Appl	186	35	41.2	641	2	US-09-199-637A-103	Sequence 103, App
114	36	42.4	614	2	US-09-248-796A-17399	Sequence 17399, A	187	35	41.2	643	2	US-09-199-637A-199	Sequence 199, App
115	36	42.4	728	2	US-08-915-337-2	Sequence 2, Appl	188	35	41.2	645	2	US-09-199-637A-253	Sequence 253, App
116	36	42.4	825	2	US-09-921-667-16	Sequence 16, Appl	189	35	41.2	646	2	US-09-252-991A-26479	Sequence 26479, A
117	36	42.4	825	2	US-09-949-001-14	Sequence 14, Appl	190	35	41.2	655	2	US-08-556-422A-3	Sequence 3, Appl
118	36	42.4	826	2	US-09-687-050-6	Sequence 6, Appl	191	35	41.2	657	2	US-09-949-016-7183	Sequence 7183, Ap
119	36	42.4	839	2	US-09-949-001-26	Sequence 26, Appl	192	35	41.2	671	2	US-09-889-746-6	Sequence 6, Appl
120	36	42.4	858	2	US-09-949-016-6702	Sequence 6702, Ap	193	35	41.2	689	1	US-08-221-817-18	Sequence 18, Appl
121	36	42.4	1209	2	US-09-949-002-493	Sequence 493, App	194	35	41.2	689	2	US-08-454-439-18	Sequence 18, Appl
122	36	42.4	1209	2	US-09-949-002-494	Sequence 494, App	195	35	41.2	689	2	US-09-771-161A-185	Sequence 185, App
123	36	42.4	2316	2	US-09-949-002-314	Sequence 314, App	196	35	41.2	689	4	PCT-US94-10487-18	Sequence 18, Appl
124	36	42.4	2415	2	US-09-949-002-398	Sequence 398, App	197	35	41.2	771	1	US-08-121-713D-54	Sequence 54, Appl
125	35.5	41.8	360	2	US-09-252-991A-31993	Sequence 31993, A	198	35	41.2	771	1	US-08-835-268-54	Sequence 54, Appl
126	35.5	41.8	807	2	US-09-081-345-2	Sequence 2, Appl	199	35	41.2	771	1	US-09-060-692-54	Sequence 54, Appl
127	35.5	41.8	807	2	US-09-822-295-2	Sequence 2, Appl	200	35	41.2	771	2	US-08-833-391-54	Sequence 54, Appl
128	35.5	41.8	1665	2	US-09-543-681A-4476	Sequence 4476, Ap	201	35	41.2	771	2	US-09-060-610-54	Sequence 54, Appl
129	35	41.2	19	1	US-08-374-652C-48	Sequence 48, Appl	202	35	41.2	771	4	PCT-US94-10151A-54	Sequence 54, Appl
130	35	41.2	51	2	US-09-270-767-60112	Sequence 60112, A	203	35	41.2	950	2	US-09-543-681A-5997	Sequence 5997, Ap
131	35	41.2	113	2	US-09-902-540-15965	Sequence 15965, A	204	35	41.2	1132	2	US-09-150-460B-11	Sequence 11, Appl
132	35	41.2	144	1	US-08-320-559-35	Sequence 35, Appl	205	35	41.2	1332	2	US-08-971-244-2	Sequence 2, Appl
133	35	41.2	144	2	US-08-545-860D-35	Sequence 35, Appl	206	35	41.2	1332	2	US-09-286-891-2	Sequence 2, Appl
134	35	41.2	144	4	PCT-US94-04496-35	Sequence 35, Appl	207	35	41.2	1349	1	US-08-612-734B-2	Sequence 2, Appl
135	35	41.2	161	2	US-09-270-767-33445	Sequence 33445, A	208	35	41.2	1523	2	US-09-538-092-955	Sequence 955, App
136	35	41.2	161	2	US-09-270-767-48662	Sequence 48662, A	209	35	41.2	1722	2	US-09-194-612A-1	Sequence 1, Appl
137	35	41.2	166	2	US-09-605-703B-2276	Sequence 2276, Ap	210	35	41.2	1722	2	US-09-949-002-341	Sequence 341, App
138	35	41.2	180	2	US-09-270-767-44659	Sequence 44659, A	211	35	41.2	1740	2	US-09-949-002-535	Sequence 535, App
139	35	41.2	181	2	US-09-603-208A-280	Sequence 280, App	212	35	41.2	1747	2	US-09-949-016-7835	Sequence 7835, Ap
140	35	41.2	188	2	US-09-107-433-3764	Sequence 3764, Ap	213	35	41.2	1927	2	US-09-949-016-6206	Sequence 6206, Ap
141	35	41.2	195	2	US-09-285-677-2	Sequence 2, Appl	214	35	41.2	2304	2	US-09-324-867-4	Sequence 4, Appl
142	35	41.2	196	1	US-08-943-208-2	Sequence 2, Appl	215	35	41.2	2319	1	US-08-212-133A-8	Sequence 8, Appl
143	35	41.2	196	2	US-09-400-861-2	Sequence 2, Appl	216	35	41.2	2319	1	US-08-474-503-6	Sequence 6, Appl
144	35	41.2	213	2	US-09-248-796A-16275	Sequence 16275, A	217	35	41.2	2319	1	US-08-670-707A-6	Sequence 6, Appl
145	35	41.2	274	2	US-09-583-110-3631	Sequence 3631, Ap	218	35	41.2	2319	2	US-09-037-601-6	Sequence 6, Appl
146	35	41.2	325	2	US-09-938-501A-2	Sequence 2, Appl	219	35	41.2	2319	2	US-09-315-179-6	Sequence 6, Appl
147	35	41.2	330	1	US-08-118-270-20	Sequence 20, Appl	220	35	41.2	2319	2	US-09-523-656-28	Sequence 28, Appl
148	35	41.2	330	4	PCT-US93-08528-20	Sequence 20, Appl	221	35	41.2	2319	4	PCT-US94-13200-6	Sequence 6, Appl
149	35	41.2	392	2	US-09-270-767-41704	Sequence 41704, A	222	34.5	40.6	30	1	US-08-287-959-13	Sequence 13, Appl
150	35	41.2	400	2	US-09-902-540-16279	Sequence 16279, A	223	34.5	40.6	253	2	US-09-605-703B-1596	Sequence 1596, Ap
151	35	41.2	419	2	US-09-485-473-2	Sequence 2, Appl	224	34.5	40.6	383	1	US-08-464-523B-31	Sequence 31, Appl
152	35	41.2	425	2	US-10-169-048-58	Sequence 58, Appl	225	34.5	40.6	425	2	US-09-634-955B-19	Sequence 19, Appl
153	35	41.2	428	2	US-09-248-796A-17801	Sequence 17801, A	226	34.5	40.6	463	2	US-09-248-796A-14184	Sequence 14184, A
154	35	41.2	430	2	US-09-770-509-26	Sequence 26, Appl	227	34.5	40.6	471	2	US-09-134-001C-4904	Sequence 4904, Ap
155	35	41.2	437	2	US-09-538-092-816	Sequence 816, App	228	34.5	40.6	501	2	US-09-949-016-11281	Sequence 11281, A
156	35	41.2	441	2	US-09-171-461-26	Sequence 26, Appl	229	34.5	40.6	1657	1	US-08-287-959-1	Sequence 1, Appl
157	35	41.2	441	2	US-09-970-711-26	Sequence 26, Appl	230	34.5	40.6	1657	2	US-09-949-016-6427	Sequence 6427, Ap
158	35	41.2	442	2	US-09-828-062-7	Sequence 7, Appl	231	34.5	40.6	1678	2	US-09-949-016-9445	Sequence 9445, Ap
159	35	41.2	445	2	US-09-991-181-177	Sequence 177, App	232	34	40.0	82	2	US-09-270-767-39198	Sequence 39198, A
160	35	41.2	445	2	US-09-990-444-177	Sequence 177, App	233	34	40.0	82	2	US-09-270-767-54415	Sequence 54415, A
161	35	41.2	445	2	US-09-997-333-177	Sequence 177, App	234	34	40.0	109	2	US-09-513-999C-7278	Sequence 7278, Ap
162	35	41.2	445	2	US-09-992-598-177	Sequence 177, App	235	34	40.0	123	2	US-09-248-796A-18357	Sequence 18357, A
163	35	41.2	447	2	US-09-825-923-2	Sequence 2, Appl	236	34	40.0	129	2	US-09-134-000C-3815	Sequence 3815, Ap
164	35	41.2	450	2	US-09-825-923-4	Sequence 4, Appl	237	34	40.0	131	2	US-10-104-047-2135	Sequence 2125, Ap
165	35	41.2	455	2	US-09-636-499-13	Sequence 13, Appl	238	34	40.0	132	2	US-09-621-976-4486	Sequence 4486, Ap
166	35	41.2	455	2	US-09-636-499-14	Sequence 14, Appl	239	34	40.0	149	2	US-09-270-767-33094	Sequence 33094, A
167	35	41.2	458	2	US-09-252-991A-30327	Sequence 30327, A	240	34	40.0	149	2	US-09-270-767-48311	Sequence 48311, A
168	35	41.2	463	2	US-09-328-352-4600	Sequence 4600, Ap	241	34	40.0	161	2	US-09-248-796A-15514	Sequence 15514, A
169	35	41.2	479	1	US-07-923-724-2	Sequence 2, Appl	242	34	40.0	172	1	US-08-468-996-7	Sequence 7, Appl
170	35	41.2	479	1	US-08-609-426A-2	Sequence 2, Appl	243	34	40.0	172	1	US-08-318-947A-11	Sequence 11, Appl
171	35	41.2	479	1	US-08-374-652C-4	Sequence 4, Appl	244	34	40.0	172	1	US-08-795-303-11	Sequence 11, Appl
172	35	41.2	484	2	US-09-605-703B-2520	Sequence 2520, Ap	245	34	40.0	172	2	US-09-358-580-14	Sequence 14, Appl
173	35	41.2	495	2	US-09-805-694B-16	Sequence 16, Appl	246	34	40.0	172	2	US-09-538-092-292	Sequence 292, App

247	34	40.0	176	2	US-09-328-352-5939	Sequence 5939, Ap	320	34	40.0	365	2	US-09-642-749-27	Sequence 27, Appl
248	34	40.0	189	2	US-10-104-047-2433	Sequence 2433, Ap	321	34	40.0	371	2	US-09-252-991A-32719	Sequence 32719, A
249	34	40.0	197	2	US-09-134-000C-6548	Sequence 6548, Ap	322	34	40.0	381	2	US-09-949-016-11291	Sequence 11291, A
250	34	40.0	202	1	US-08-855-261A-4	Sequence 4, Appli	323	34	40.0	387	2	US-09-252-991A-17881	Sequence 17881, A
251	34	40.0	202	1	US-08-839-711-4	Sequence 4, Appli	324	34	40.0	390	2	US-09-419-459-4	Sequence 4, Appli
252	34	40.0	202	2	US-09-227-224-4	Sequence 4, Appli	325	34	40.0	391	2	US-09-248-796A-20440	Sequence 20440, A
253	34	40.0	202	2	US-09-855-288-10	Sequence 10, Appl	326	34	40.0	394	2	US-09-419-459-10	Sequence 10, Appl
254	34	40.0	212	2	US-08-811-463-39	Sequence 39, Appl	327	34	40.0	409	2	US-10-332-795-17	Sequence 17, Appl
255	34	40.0	212	2	US-09-933-497B-39	Sequence 39, Appl	328	34	40.0	411	2	US-09-015-188-2	Sequence 2, Appli
256	34	40.0	213	2	US-09-231-170A-5	Sequence 5, Appli	329	34	40.0	424	2	US-09-252-991A-23630	Sequence 23630, A
257	34	40.0	213	2	US-09-724-884-5	Sequence 5, Appli	330	34	40.0	438	2	US-09-248-796A-15113	Sequence 15113, A
258	34	40.0	213	2	US-09-724-592-5	Sequence 5, Appli	331	34	40.0	440	2	US-09-457-046B-26	Sequence 26, Appl
259	34	40.0	213	2	US-09-673-222-5	Sequence 5, Appli	332	34	40.0	440	2	US-09-866-570B-26	Sequence 26, Appl
260	34	40.0	215	2	US-09-483-588-8	Sequence 8, Appli	333	34	40.0	445	1	US-08-353-400-33	Sequence 33, Appl
261	34	40.0	221	2	US-09-107-532A-4477	Sequence 4477, Ap	334	34	40.0	451	2	US-09-107-532A-6444	Sequence 6444, Ap
262	34	40.0	221	2	US-09-134-000C-6523	Sequence 6523, Ap	335	34	40.0	457	2	US-09-270-767-42082	Sequence 42082, A
263	34	40.0	227	2	US-09-673-245-5	Sequence 5, Appli	336	34	40.0	464	1	US-08-353-400-36	Sequence 36, Appl
264	34	40.0	236	2	US-09-489-039A-13791	Sequence 13791, A	337	34	40.0	479	2	US-09-949-016-6869	Sequence 6869, Ap
265	34	40.0	237	2	US-09-489-039A-8912	Sequence 8912, Ap	338	34	40.0	485	2	US-09-252-991A-17170	Sequence 17170, A
266	34	40.0	241	2	US-08-896-933-32	Sequence 32, Appl	339	34	40.0	490	2	US-09-949-016-7057	Sequence 7057, Ap
267	34	40.0	241	2	US-09-314-235-32	Sequence 32, Appl	340	34	40.0	503	1	US-07-946-497-2	Sequence 2, Appli
268	34	40.0	241	2	US-09-708-008B-32	Sequence 32, Appl	341	34	40.0	503	1	US-08-483-322-2	Sequence 2, Appli
269	34	40.0	243	1	US-08-460-309-16	Sequence 16, Appl	342	34	40.0	503	1	US-08-478-882-2	Sequence 2, Appli
270	34	40.0	243	1	US-08-125-077-16	Sequence 16, Appl	343	34	40.0	525	2	US-09-248-796A-19506	Sequence 19506, A
271	34	40.0	250	2	US-09-653-778B-2	Sequence 2, Appli	344	34	40.0	528	2	US-09-149-476-732	Sequence 732, App
272	34	40.0	263	2	US-09-248-796A-14226	Sequence 14226, A	345	34	40.0	548	2	US-09-149-476-469	Sequence 469, App
273	34	40.0	264	2	US-09-107-532A-4481	Sequence 4481, Ap	346	34	40.0	549	2	US-09-107-532A-6840	Sequence 6840, Ap
274	34	40.0	265	2	US-09-538-092-164	Sequence 164, App	347	34	40.0	558	2	US-09-513-783A-180	Sequence 180, App
275	34	40.0	270	2	US-09-362-473-10	Sequence 10, Appl	348	34	40.0	558	2	US-10-100-957A-180	Sequence 180, App
276	34	40.0	286	2	US-09-489-039A-11182	Sequence 11182, A	349	34	40.0	562	2	US-09-673-245-7	Sequence 7, Appli
277	34	40.0	286	2	US-09-270-767-44446	Sequence 44446, A	350	34	40.0	570	2	US-09-949-016-8907	Sequence 8907, Ap
278	34	40.0	304	2	US-09-107-532A-5424	Sequence 5424, Ap	351	34	40.0	573	1	US-08-991-531-1	Sequence 1, Appli
279	34	40.0	306	2	US-09-457-046B-4	Sequence 4, Appli	352	34	40.0	573	1	US-09-032-315-10	Sequence 10, Appl
280	34	40.0	306	2	US-09-866-570B-4	Sequence 4, Appli	353	34	40.0	573	1	US-08-993-318A-10	Sequence 10, Appl
281	34	40.0	314	1	US-08-989-478-3	Sequence 3, Appli	354	34	40.0	573	2	US-09-028-887-1	Sequence 1, Appli
282	34	40.0	314	1	US-08-989-478-4	Sequence 4, Appli	355	34	40.0	573	2	US-09-399-886-10	Sequence 10, Appl
283	34	40.0	314	1	US-08-989-478-5	Sequence 5, Appli	356	34	40.0	573	2	US-09-396-260-10	Sequence 10, Appl
284	34	40.0	314	2	US-08-996-685-3	Sequence 3, Appli	357	34	40.0	573	2	US-09-518-901-1	Sequence 1, Appli
285	34	40.0	314	2	US-08-996-685-4	Sequence 4, Appli	358	34	40.0	573	2	US-09-576-281-10	Sequence 10, Appl
286	34	40.0	314	2	US-08-996-685-5	Sequence 5, Appli	359	34	40.0	577	2	US-10-104-047-2569	Sequence 2569, Ap
287	34	40.0	317	1	US-08-466-337A-18	Sequence 18, Appl	360	34	40.0	580	2	US-09-252-991A-29629	Sequence 29629, A
288	34	40.0	317	1	US-08-475-359-18	Sequence 18, Appl	361	34	40.0	581	2	US-09-270-767-45723	Sequence 45723, A
289	34	40.0	317	1	US-08-802-322-1	Sequence 1, Appli	362	34	40.0	582	2	US-09-419-459-2	Sequence 2, Appli
290	34	40.0	317	2	US-08-465-887A-18	Sequence 18, Appl	363	34	40.0	599	1	US-08-442-542-18	Sequence 18, Appl
291	34	40.0	317	2	US-08-895-601-7	Sequence 7, Appli	364	34	40.0	599	2	US-08-765-469-18	Sequence 18, Appl
292	34	40.0	317	2	US-09-349-627-1	Sequence 1, Appli	365	34	40.0	601	2	US-09-889-463A-36	Sequence 36, Appl
293	34	40.0	317	2	US-08-910-820-1	Sequence 1, Appli	366	34	40.0	620	1	US-08-706-037-27	Sequence 27, Appl
294	34	40.0	317	2	US-09-844-908-1	Sequence 1, Appli	367	34	40.0	620	1	US-08-940-661A-2	Sequence 2, Appli
295	34	40.0	317	2	US-09-832-161-1	Sequence 1, Appli	368	34	40.0	620	1	US-09-083-485-2	Sequence 2, Appli
296	34	40.0	317	2	US-10-052-005A-9	Sequence 9, Appli	369	34	40.0	620	1	US-09-005-397-27	Sequence 27, Appl
297	34	40.0	317	2	US-09-949-016-6969	Sequence 6969, Ap	370	34	40.0	620	1	US-08-939-218A-2	Sequence 2, Appli
298	34	40.0	318	2	US-09-489-039A-13767	Sequence 13767, A	371	34	40.0	620	4	PCT-US95-06815-2	Sequence 2, Appli
299	34	40.0	324	1	US-08-579-940-7	Sequence 7, Appli	372	34	40.0	627	2	US-09-303-518D-456	Sequence 456, App
300	34	40.0	334	1	US-08-359-850-4	Sequence 4, Appli	373	34	40.0	773	1	US-08-484-101B-42	Sequence 42, Appl
301	34	40.0	339	2	US-09-270-767-32823	Sequence 32823, A	374	34	40.0	773	1	US-08-484-101B-42	Sequence 42, Appl
302	34	40.0	339	2	US-09-270-767-48040	Sequence 48040, A	375	34	40.0	773	2	US-08-714-524D-44	Sequence 44, Appl
303	34	40.0	347	1	US-08-802-322-12	Sequence 12, Appl	376	34	40.0	773	2	US-08-714-524D-44	Sequence 44, Appl
304	34	40.0	347	1	US-08-802-322-13	Sequence 13, Appl	377	34	40.0	809	1	US-07-789-915A-4	Sequence 4, Appli
305	34	40.0	347	2	US-09-832-161-12	Sequence 12, Appl	378	34	40.0	809	1	US-08-005-002C-4	Sequence 4, Appli
306	34	40.0	347	2	US-09-832-161-13	Sequence 13, Appl	379	34	40.0	809	1	US-08-487-203A-4	Sequence 4, Appli
307	34	40.0	348	2	US-09-949-016-7384	Sequence 7384, Ap	380	34	40.0	815	2	US-09-339-159B-26	Sequence 26, Appl
308	34	40.0	355	1	US-07-946-497-5	Sequence 5, Appli	381	34	40.0	839	2	US-09-902-540-15096	Sequence 15096, A
309	34	40.0	355	1	US-08-483-322-5	Sequence 5, Appli	382	34	40.0	850	2	US-09-328-352-7660	Sequence 7660, Ap
310	34	40.0	355	1	US-08-478-882-5	Sequence 5, Appli	383	34	40.0	901	2	US-09-134-001C-5351	Sequence 5351, Ap
311	34	40.0	355	1	US-08-674-612-2	Sequence 2, Appli	384	34	40.0	923	2	US-09-540-236-3837	Sequence 3837, Ap
312	34	40.0	355	1	US-08-920-296-2	Sequence 2, Appli	385	34	40.0	925	2	US-09-924-097A-14	Sequence 14, Appl
313	34	40.0	365	1	US-08-746-788-2	Sequence 2, Appli	386	34	40.0	928	2	US-09-252-991A-24200	Sequence 24200, A
314	34	40.0	365	1	US-09-189-602-2	Sequence 2, Appli	387	34	40.0	1036	2	US-09-489-039A-9371	Sequence 9371, Ap
315	34	40.0	365	2	US-09-025-580-26	Sequence 26, Appl	388	34	40.0	1061	2	US-10-200-012-4	Sequence 4, Appli
316	34	40.0	365	2	US-09-025-580-27	Sequence 27, Appl	389	34	40.0	1265	2	US-09-347-878-5	Sequence 5, Appli
317	34	40.0	365	2	US-09-124-163-2	Sequence 2, Appli	390	34	40.0	1265	2	US-09-347-878-7	Sequence 7, Appli
318	34	40.0	365	2	US-09-809-464-1	Sequence 1, Appli	391	34	40.0	1265	2	US-08-980-326-2	Sequence 2, Appli
319	34	40.0	365	2	US-09-642-749-26	Sequence 26, Appl	392	34	40.0	1265	2	US-08-980-326-74	Sequence 74, Appl

393	33.5	39.4	50	2	US-08-975-080-26	Sequence 26, Appl	466	33	38.8	138	2	US-09-489-039A-7717	Sequence 7717, Ap
394	33.5	39.4	50	2	US-10-138-618-26	Sequence 26, Appl	467	33	38.8	199	2	US-09-543-681A-4701	Sequence 4701, Ap
395	33.5	39.4	50	2	US-09-690-825-26	Sequence 26, Appl	468	33	38.8	203	2	US-09-270-767-42051	Sequence 42051, A
396	33.5	39.4	56	2	US-08-657-759-24	Sequence 24, Appl	469	33	38.8	207	2	US-09-248-796A-14709	Sequence 14709, A
397	33.5	39.4	109	2	US-09-248-796A-27688	Sequence 27688, A	470	33	38.8	220	2	US-09-252-991A-30617	Sequence 30617, A
398	33.5	39.4	127	2	US-09-513-999C-7853	Sequence 7853, Ap	471	33	38.8	229	2	US-09-248-796A-15649	Sequence 15649, A
399	33.5	39.4	241	2	US-08-936-165A-362	Sequence 362, App	472	33	38.8	258	2	US-09-252-991A-32929	Sequence 32929, A
400	33.5	39.4	261	2	US-09-902-540-13754	Sequence 13754, A	473	33	38.8	271	2	US-09-355-166-10	Sequence 10, Appl
401	33.5	39.4	359	2	US-09-248-796A-25757	Sequence 25757, A	474	33	38.8	277	2	US-09-949-016-6212	Sequence 6212, Ap
402	33.5	39.4	403	2	US-09-248-796A-14525	Sequence 14525, A	475	33	38.8	284	2	US-09-952-689-3	Sequence 3, Appli
403	33.5	39.4	517	2	US-09-902-540-16073	Sequence 16073, A	476	33	38.8	287	2	US-09-949-016-8068	Sequence 8068, Ap
404	33.5	39.4	723	2	US-09-949-016-9810	Sequence 9810, Ap	477	33	38.8	291	2	US-09-315-793-62	Sequence 62, Appl
405	33.5	39.4	1076	2	US-09-976-594-889	Sequence 889, App	478	33	38.8	291	2	US-09-270-767-32678	Sequence 32678, A
406	33.5	39.4	1151	2	US-08-836-134-23	Sequence 23, Appl	479	33	38.8	291	2	US-09-270-767-47895	Sequence 47895, A
407	33.5	39.4	1151	2	US-08-493-784-23	Sequence 23, Appl	480	33	38.8	291	2	US-09-307-621-4	Sequence 4, Appli
408	33.5	39.4	1194	2	US-09-092-508-2	Sequence 2, Appli	481	33	38.8	294	2	US-09-134-000C-4724	Sequence 4724, Ap
409	33.5	39.4	1194	2	US-09-435-115-2	Sequence 2, Appli	482	33	38.8	298	2	US-09-248-796A-17360	Sequence 17360, A
410	33.5	39.4	1194	2	US-09-069-123-26	Sequence 26, Appl	483	33	38.8	300	2	US-09-198-452A-426	Sequence 426, App
411	33.5	39.4	1194	2	US-09-098-310-2	Sequence 2, Appli	484	33	38.8	300	2	US-09-248-796A-15243	Sequence 15243, A
412	33.5	39.4	1194	2	US-09-538-092-825	Sequence 825, App	485	33	38.8	300	2	US-09-438-185A-409	Sequence 409, App
413	33.5	39.4	1194	2	US-09-949-016-6030	Sequence 6030, Ap	486	33	38.8	311	2	US-09-489-039A-11573	Sequence 11573, A
414	33.5	39.4	1196	2	US-09-949-016-10065	Sequence 10065, A	487	33	38.8	315	2	US-09-248-796A-15110	Sequence 15110, A
415	33.5	39.4	1196	2	US-09-949-016-10066	Sequence 10066, A	488	33	38.8	324	2	US-09-107-532A-4762	Sequence 4762, Ap
416	33.5	39.4	1205	2	US-09-092-508-16	Sequence 16, Appl	489	33	38.8	338	2	US-09-248-796A-24205	Sequence 24205, A
417	33.5	39.4	1205	2	US-09-435-115-16	Sequence 16, Appl	490	33	38.8	340	2	US-09-252-991A-20826	Sequence 20826, A
418	33.5	39.4	1232	2	US-08-836-134-2	Sequence 2, Appli	491	33	38.8	351	2	US-09-443-067-18	Sequence 18, Appl
419	33.5	39.4	1232	2	US-09-493-784-2	Sequence 2, Appli	492	33	38.8	357	2	US-09-605-703B-2060	Sequence 2060, Ap
420	33.5	39.4	1237	2	US-09-949-016-6842	Sequence 6842, Ap	493	33	38.8	363	2	US-09-041-718-5	Sequence 5, Appli
421	33.5	39.4	1239	2	US-09-949-016-10063	Sequence 10063, A	494	33	38.8	363	2	US-09-976-594-913	Sequence 913, App
422	33.5	39.4	1239	2	US-09-949-016-10064	Sequence 10064, A	495	33	38.8	380	2	US-09-248-796A-16841	Sequence 16841, A
423	33.5	39.4	1295	2	US-09-705-872-3	Sequence 3, Appli	496	33	38.8	385	2	US-09-543-681A-7438	Sequence 7438, Ap
424	33.5	39.4	1301	2	US-09-252-991A-29822	Sequence 29822, A	497	33	38.8	385	2	US-09-248-796A-19434	Sequence 19434, A
425	33.5	39.4	1403	2	US-09-705-872-1	Sequence 1, Appli	498	33	38.8	398	2	US-09-248-796A-25567	Sequence 25567, A
426	33.5	39.4	1403	2	US-09-949-002-370	Sequence 370, App	499	33	38.8	412	1	US-08-313-288B-18	Sequence 18, Appl
427	33.5	39.4	1412	2	US-09-949-002-411	Sequence 411, App	500	33	38.8	418	2	US-09-134-001C-3753	Sequence 3753, Ap
428	33	38.8	20	1	US-08-162-149-10	Sequence 10, Appl	501	33	38.8	418	2	US-09-489-039A-11946	Sequence 11946, A
429	33	38.8	20	2	US-08-264-578-3	Sequence 3, Appli	502	33	38.8	423	1	US-08-760-797A-1	Sequence 1, Appli
430	33	38.8	20	2	US-08-264-578-6	Sequence 6, Appli	503	33	38.8	424	1	US-08-760-797A-3	Sequence 3, Appli
431	33	38.8	30	1	US-08-588-873-46	Sequence 46, Appl	504	33	38.8	424	2	US-08-932-929B-1	Sequence 1, Appli
432	33	38.8	30	2	US-08-605-430-46	Sequence 46, Appl	505	33	38.8	424	2	US-08-932-929B-3	Sequence 3, Appli
433	33	38.8	30	2	US-09-717-054-46	Sequence 46, Appl	506	33	38.8	428	1	US-08-907-706-5	Sequence 5, Appli
434	33	38.8	51	2	US-09-513-999C-5287	Sequence 5287, Ap	507	33	38.8	430	2	US-09-198-452A-497	Sequence 497, App
435	33	38.8	65	2	US-09-489-039A-12573	Sequence 12573, A	508	33	38.8	432	2	US-09-438-185A-465	Sequence 465, App
436	33	38.8	65	2	US-09-248-796A-27138	Sequence 27138, A	509	33	38.8	432	2	US-09-491-577-94	Sequence 94, Appl
437	33	38.8	83	2	US-09-621-976-5869	Sequence 5869, Ap	510	33	38.8	436	2	US-09-252-991A-22426	Sequence 22426, A
438	33	38.8	87	2	US-09-344-624-15	Sequence 15, Appl	511	33	38.8	444	2	US-09-252-991A-32415	Sequence 32415, A
439	33	38.8	109	2	US-09-270-767-40598	Sequence 40598, A	512	33	38.8	453	2	US-09-252-991A-31268	Sequence 31268, A
440	33	38.8	109	2	US-09-270-767-55814	Sequence 55814, A	513	33	38.8	461	2	US-09-949-016-8508	Sequence 8508, Ap
441	33	38.8	114	2	US-08-545-809A-124	Sequence 124, App	514	33	38.8	463	1	US-08-907-706-1	Sequence 1, Appli
442	33	38.8	114	2	US-09-621-976-5693	Sequence 5693, Ap	515	33	38.8	467	2	US-09-573-395A-308	Sequence 308, App
443	33	38.8	114	2	US-09-515-697-124	Sequence 124, App	516	33	38.8	490	2	US-09-252-991A-30874	Sequence 30874, A
444	33	38.8	125	2	US-09-248-796A-14710	Sequence 14710, A	517	33	38.8	501	2	US-09-328-352-6940	Sequence 6940, Ap
445	33	38.8	131	2	US-09-248-796A-23977	Sequence 23977, A	518	33	38.8	503	2	US-09-562-737-69	Sequence 69, Appl
446	33	38.8	146	1	US-08-173-510B-88	Sequence 88, Appl	519	33	38.8	522	2	US-09-710-279-604	Sequence 604, App
447	33	38.8	146	1	US-08-458-218-86	Sequence 86, Appl	520	33	38.8	525	2	US-09-252-991A-19459	Sequence 19459, A
448	33	38.8	146	1	US-08-450-497-88	Sequence 88, Appl	521	33	38.8	526	2	US-09-447-497-15	Sequence 15, Appl
449	33	38.8	146	2	US-08-060-433C-38	Sequence 38, Appl	522	33	38.8	531	2	US-09-922-011-11	Sequence 11, Appl
450	33	38.8	146	2	US-08-450-482B-88	Sequence 88, Appl	523	33	38.8	542	2	US-10-037-417-97	Sequence 97, Appl
451	33	38.8	146	2	US-08-151-064D-86	Sequence 86, Appl	524	33	38.8	544	1	US-08-264-002-7	Sequence 7, Appli
452	33	38.8	148	1	US-08-392-625-18	Sequence 18, Appl	525	33	38.8	544	2	US-09-457-040B-15	Sequence 15, Appl
453	33	38.8	148	1	US-08-466-961A-18	Sequence 18, Appl	526	33	38.8	553	2	US-09-447-497-14	Sequence 14, Appl
454	33	38.8	148	1	US-08-645-193B-13	Sequence 13, Appl	527	33	38.8	553	2	US-10-104-047-2655	Sequence 2655, Ap
455	33	38.8	149	2	US-09-252-991A-24978	Sequence 24978, A	528	33	38.8	564	2	US-09-252-991A-23909	Sequence 23909, A
456	33	38.8	153	2	US-09-902-540-13661	Sequence 13661, A	529	33	38.8	576	2	US-09-248-796A-20509	Sequence 20509, A
457	33	38.8	162	2	US-09-270-767-33361	Sequence 33361, A	530	33	38.8	588	1	US-09-005-232A-2	Sequence 2, Appli
458	33	38.8	166	2	US-09-248-796A-14520	Sequence 14520, A	531	33	38.8	588	1	US-09-005-232A-2	Sequence 2, Appli
459	33	38.8	169	2	US-09-252-991A-28109	Sequence 28109, A	532	33	38.8	592	2	US-08-961-083-222	Sequence 222, App
460	33	38.8	173	2	US-09-270-767-34380	Sequence 34380, A	533	33	38.8	592	2	US-09-536-784-222	Sequence 222, App
461	33	38.8	174	2	US-09-270-767-49597	Sequence 49597, A	534	33	38.8	592	2	US-09-765-271-222	Sequence 222, App
462	33	38.8	174	2	US-09-270-767-2612	Sequence 2612, Ap	535	33	38.8	605	2	US-09-765-272A-222	Sequence 222, App
463	33	38.8	186	2	US-09-248-796A-15844	Sequence 15844, A	536	33	38.8	608	2	US-09-270-767-46196	Sequence 46196, A
464	33	38.8	196	2	US-09-842-164A-2	Sequence 2, Appli	537	33	38.8	608	2	US-09-447-497-6	Sequence 6, Appli
465	33	38.8	197	2	US-09-112-248-2	Sequence 2, Appli	538	33	38.8	609	2	US-09-447-497-12	Sequence 12, Appl

539	33	38.8	612	2	US-09-447-497-3	Sequence 3, Appli	612	33	38.8	2349	2	US-09-538-092-914	Sequence 914, App
540	33	38.8	613	2	US-09-447-497-9	Sequence 9, Appli	613	33	38.8	4472	1	US-08-804-227C-2	Sequence 2, Appli
541	33	38.8	614	2	US-09-806-536A-9	Sequence 9, Appli	614	33	38.8	4545	1	US-08-804-227C-14	Sequence 14, Appli
542	33	38.8	620	2	US-09-673-198-1	Sequence 1, Appli	615	32.5	38.2	172	2	US-08-858-207A-477	Sequence 477, App
543	33	38.8	623	2	US-09-447-497-11	Sequence 11, Appli	616	32.5	38.2	289	2	US-09-248-796A-19422	Sequence 19422, A
544	33	38.8	625	2	US-09-447-497-5	Sequence 5, Appli	617	32.5	38.2	295	2	US-09-583-110-2917	Sequence 2917, Ap
545	33	38.8	626	2	US-09-489-039A-13113	Sequence 13113, A	618	32.5	38.2	307	2	US-09-197-970B-3	Sequence 3, Appli
546	33	38.8	627	2	US-09-447-497-8	Sequence 8, Appli	619	32.5	38.2	307	2	US-09-583-110-4520	Sequence 4520, Ap
547	33	38.8	627	2	US-09-949-016-7354	Sequence 7354, Ap	620	32.5	38.2	313	2	US-09-107-433-4450	Sequence 4450, Ap
548	33	38.8	629	2	US-09-447-497-2	Sequence 2, Appli	621	32.5	38.2	327	2	US-09-107-433-2743	Sequence 2743, Ap
549	33	38.8	636	2	US-09-489-039A-8201	Sequence 8201, Ap	622	32.5	38.2	333	2	US-09-769-787-43	Sequence 43, Appli
550	33	38.8	642	2	US-09-949-016-10283	Sequence 10283, A	623	32.5	38.2	341	2	US-09-248-796A-20844	Sequence 20844, A
551	33	38.8	666	2	US-09-158-452A-409	Sequence 409, App	624	32.5	38.2	347	2	US-09-252-991A-25606	Sequence 25606, A
552	33	38.8	666	2	US-09-438-185A-390	Sequence 390, App	625	32.5	38.2	361	2	US-09-270-767-42397	Sequence 42397, A
553	33	38.8	674	2	US-09-270-767-41618	Sequence 41618, A	626	32.5	38.2	363	2	US-09-252-991A-18648	Sequence 18648, A
554	33	38.8	679	2	US-09-270-767-44535	Sequence 44535, A	627	32.5	38.2	490	1	US-08-294-770A-2	Sequence 2, Appli
555	33	38.8	690	2	US-08-971-089-6	Sequence 6, Appli	628	32.5	38.2	491	1	US-08-448-735C-2	Sequence 1, Appli
556	33	38.8	690	2	US-10-117-604A-6	Sequence 6, Appli	629	32.5	38.2	491	2	US-09-958-561A-1	Sequence 1, Appli
557	33	38.8	755	2	US-09-949-016-7755	Sequence 7755, Ap	630	32.5	38.2	501	2	US-09-248-796A-16289	Sequence 16289, A
558	33	38.8	760	2	US-09-328-352-7293	Sequence 7293, Ap	631	32.5	38.2	534	2	US-09-328-352-6713	Sequence 6713, Ap
559	33	38.8	766	2	US-09-252-991A-31826	Sequence 31826, A	632	32.5	38.2	575	2	US-09-248-796A-17507	Sequence 17507, A
560	33	38.8	775	2	US-09-328-352-7095	Sequence 7095, Ap	633	32.5	38.2	765	2	US-09-248-796A-20068	Sequence 20068, A
561	33	38.8	776	2	US-09-266-225D-10	Sequence 10, Appli	634	32.5	38.2	979	2	US-09-543-681A-5466	Sequence 5466, Ap
562	33	38.8	777	2	US-09-949-016-6540	Sequence 6540, Ap	635	32.5	38.2	1242	2	US-09-540-236-2522	Sequence 2522, Ap
563	33	38.8	803	2	US-09-489-039A-12742	Sequence 12742, A	636	32.5	38.2	1577	1	US-08-793-82A-2	Sequence 2, Appli
564	33	38.8	805	2	US-09-344-624-19	Sequence 19, Appli	637	32	37.6	21	1	US-08-127-499A-18	Sequence 18, Appli
565	33	38.8	805	2	US-09-949-016-6713	Sequence 6713, Ap	638	32	37.6	21	1	US-08-482-847-18	Sequence 18, Appli
566	33	38.8	870	2	US-09-177-650-91	Sequence 91, Appli	639	32	37.6	33	2	US-09-270-767-57805	Sequence 57805, A
567	33	38.8	886	2	US-10-002-309B-2	Sequence 2, Appli	640	32	37.6	38	1	US-08-176-500-78	Sequence 78, Appli
568	33	38.8	887	1	US-08-066-167-3	Sequence 3, Appli	641	32	37.6	38	1	US-08-471-052A-78	Sequence 78, Appli
569	33	38.8	911	1	US-08-487-890A-95	Sequence 95, Appli	642	32	37.6	38	1	US-08-189-331-78	Sequence 78, Appli
570	33	38.8	911	1	US-08-478-435-95	Sequence 95, Appli	643	32	37.6	38	1	US-08-471-939-78	Sequence 78, Appli
571	33	38.8	911	1	US-08-337-483-95	Sequence 95, Appli	644	32	37.6	38	1	US-08-471-800-78	Sequence 78, Appli
572	33	38.8	911	1	US-08-478-373-95	Sequence 95, Appli	645	32	37.6	38	1	US-08-471-068-78	Sequence 78, Appli
573	33	38.8	911	2	US-08-474-671-95	Sequence 95, Appli	646	32	37.6	60	2	US-09-513-999C-6612	Sequence 6612, Ap
574	33	38.8	911	2	US-08-483-577A-95	Sequence 95, Appli	647	32	37.6	62	2	US-09-328-352-7395	Sequence 7395, Ap
575	33	38.8	911	2	US-08-448-194-6	Sequence 6, Appli	648	32	37.6	63	2	US-09-583-110-5256	Sequence 5256, Ap
576	33	38.8	911	2	US-08-613-009A-17	Sequence 17, Appli	649	32	37.6	63	2	US-09-107-433-3673	Sequence 3673, Ap
577	33	38.8	911	2	US-08-897-438-95	Sequence 95, Appli	650	32	37.6	64	2	US-09-643-657-6	Sequence 6, Appli
578	33	38.8	911	2	US-08-867-921-6	Sequence 6, Appli	651	32	37.6	72	2	US-09-621-971A-7585	Sequence 7585, Ap
579	33	38.8	911	2	US-08-637-654-95	Sequence 95, Appli	652	32	37.6	73	2	US-09-543-681A-6595	Sequence 6595, Ap
580	33	38.8	911	2	US-08-649-518-95	Sequence 95, Appli	653	32	37.6	75	2	US-09-270-767-36514	Sequence 36514, A
581	33	38.8	911	2	US-08-778-570B-23	Sequence 23, Appli	654	32	37.6	75	2	US-09-270-767-51731	Sequence 51731, A
582	33	38.8	911	2	US-09-059-584-23	Sequence 23, Appli	655	32	37.6	79	2	US-09-248-796A-24793	Sequence 24793, A
583	33	38.8	1131	2	US-09-487-558B-72	Sequence 72, Appli	656	32	37.6	91	2	US-09-270-767-42508	Sequence 42508, A
584	33	38.8	1233	2	US-09-538-092-777	Sequence 77, App	657	32	37.6	99	1	US-08-341-843B-26	Sequence 26, Appli
585	33	38.8	1237	1	US-08-241-853-2	Sequence 2, Appli	658	32	37.6	99	1	US-08-427-497B-31	Sequence 31, Appli
586	33	38.8	1237	1	US-08-850-917-2	Sequence 2, Appli	659	32	37.6	99	2	US-09-248-796A-14986	Sequence 14986, A
587	33	38.8	1310	2	US-10-373-090A-1	Sequence 1, Appli	660	32	37.6	102	2	US-09-270-767-44517	Sequence 44517, A
588	33	38.8	1347	2	US-09-949-016-9603	Sequence 9603, Ap	661	32	37.6	109	2	US-09-482-273-171	Sequence 171, App
589	33	38.8	1401	2	US-08-781-891-206	Sequence 206, App	662	32	37.6	111	2	US-09-328-352-6450	Sequence 6450, Ap
590	33	38.8	1401	2	US-09-127-670-6	Sequence 6, Appli	663	32	37.6	113	1	US-08-248-839C-2	Sequence 2, Appli
591	33	38.8	1401	2	US-09-618-166-206	Sequence 206, App	664	32	37.6	113	1	US-08-248-839C-4	Sequence 4, Appli
592	33	38.8	1456	1	US-08-026-138B-8	Sequence 8, Appli	665	32	37.6	113	1	US-08-248-839C-6	Sequence 6, Appli
593	33	38.8	1480	2	US-09-922-011-10	Sequence 10, Appli	666	32	37.6	113	1	US-08-248-839C-8	Sequence 8, Appli
594	33	38.8	1482	1	US-08-026-138B-2	Sequence 2, Appli	667	32	37.6	114	2	US-09-248-796A-27842	Sequence 27842, A
595	33	38.8	1484	1	US-08-231-193A-56	Sequence 56, Appli	668	32	37.6	116	2	US-09-252-991A-32519	Sequence 32519, A
596	33	38.8	1484	1	US-08-486-273A-56	Sequence 56, Appli	669	32	37.6	119	2	US-09-482-273-264	Sequence 264, App
597	33	38.8	1484	2	US-08-940-086A-56	Sequence 56, Appli	670	32	37.6	121	2	US-09-270-767-36220	Sequence 36220, A
598	33	38.8	1484	2	US-08-940-035A-56	Sequence 56, Appli	671	32	37.6	121	2	US-09-270-767-51437	Sequence 51437, A
599	33	38.8	1484	2	US-08-935-105A-56	Sequence 56, Appli	672	32	37.6	124	2	US-09-513-999C-6283	Sequence 6283, Ap
600	33	38.8	1484	2	US-08-264-578-2	Sequence 2, Appli	673	32	37.6	128	2	US-10-104-047-3661	Sequence 3661, Ap
601	33	38.8	1484	2	US-09-648-797-56	Sequence 56, Appli	674	32	37.6	133	2	US-09-900-920-58	Sequence 58, Appli
602	33	38.8	1484	2	US-09-386-123-56	Sequence 56, Appli	675	32	37.6	135	2	US-09-513-999C-4637	Sequence 4637, Ap
603	33	38.8	1484	2	US-09-949-016-5998	Sequence 5998, Ap	676	32	37.6	143	2	US-08-946-329A-52	Sequence 52, Appli
604	33	38.8	1484	2	US-10-038-337-56	Sequence 56, Appli	677	32	37.6	143	2	US-09-562-914-52	Sequence 52, Appli
605	33	38.8	1484	2	US-10-007-747-56	Sequence 56, Appli	678	32	37.6	147	2	US-09-640-211A-2251	Sequence 2251, Ap
606	33	38.8	1484	2	US-09-945-901-56	Sequence 56, Appli	679	32	37.6	151	2	US-09-252-991A-18270	Sequence 18270, A
607	33	38.8	1728	2	US-09-949-002-532	Sequence 532, App	680	32	37.6	157	2	US-09-270-767-42022	Sequence 42022, A
608	33	38.8	1829	2	US-09-157-420-1	Sequence 1, Appli	681	32	37.6	161	2	US-09-134-001C-4022	Sequence 4022, Ap
609	33	38.8	1963	2	US-09-583-110-5243	Sequence 5243, Ap	682	32	37.6	162	2	US-09-248-796A-21408	Sequence 21408, A
610	33	38.8	1967	2	US-09-107-433-4883	Sequence 4883, Ap	683	32	37.6	163	1	US-09-066-074-2	Sequence 2, Appli
611	33	38.8	2186	2	US-09-949-016-10828	Sequence 10828, A	684	32	37.6	163	1	US-08-555-912A-2	Sequence 2, Appli



685	32	37.6	163	2	US-09-208-804-4	Sequence 4, Appli	758	32	37.6	348	2	US-10-010-160-14	Sequence 14, Appli
686	32	37.6	163	2	US-08-801-743-4	Sequence 4, Appli	759	32	37.6	351	2	US-08-817-441-47	Sequence 47, Appli
687	32	37.6	163	2	US-09-275-900-2	Sequence 2, Appli	760	32	37.6	354	2	US-09-107-532A-7233	Sequence 7233, Ap
688	32	37.6	168	2	US-09-949-016-8592	Sequence 8592, Ap	761	32	37.6	355	2	US-09-949-016-10717	Sequence 10717, A
689	32	37.6	169	2	US-09-134-001C-2948	Sequence 2948, Ap	762	32	37.6	361	2	US-09-790-179-2	Sequence 2, Appli
690	32	37.6	170	2	US-08-468-996-9	Sequence 9, Appli	763	32	37.6	361	2	US-10-165-800-2	Sequence 2, Appli
691	32	37.6	177	2	US-09-270-767-32541	Sequence 32541, A	764	32	37.6	362	2	US-09-485-648-6	Sequence 6, Appli
692	32	37.6	177	2	US-09-270-767-47758	Sequence 47758, A	765	32	37.6	362	2	US-09-503-565-6	Sequence 6, Appli
693	32	37.6	181	2	US-09-602-787A-40	Sequence 40, Appli	766	32	37.6	362	2	US-09-485-649-6	Sequence 6, Appli
694	32	37.6	185	2	US-09-248-796A-25259	Sequence 25259, A	767	32	37.6	365	2	US-09-602-777A-394	Sequence 394, App
695	32	37.6	190	1	US-08-924-759-16	Sequence 16, Appli	768	32	37.6	368	2	US-09-902-540-12883	Sequence 12883, A
696	32	37.6	190	2	US-09-248-335-16	Sequence 16, Appli	769	32	37.6	386	2	US-09-134-001C-5607	Sequence 5607, Ap
697	32	37.6	190	2	US-09-634-238-350	Sequence 350, App	770	32	37.6	389	2	US-09-252-991A-18644	Sequence 18644, A
698	32	37.6	192	2	US-09-792-024-116	Sequence 116, App	771	32	37.6	391	2	US-09-583-110-3765	Sequence 3765, Ap
699	32	37.6	197	2	US-08-997-020-7	Sequence 7, Appli	772	32	37.6	391	2	US-09-248-796A-16042	Sequence 16042, A
700	32	37.6	197	2	US-09-350-823-7	Sequence 7, Appli	773	32	37.6	396	2	US-09-252-991A-19425	Sequence 19425, A
701	32	37.6	204	2	US-08-506-296B-58	Sequence 58, Appli	774	32	37.6	399	2	US-09-222-938A-49	Sequence 49, Appli
702	32	37.6	205	2	US-10-172-527A-30	Sequence 30, Appli	775	32	37.6	403	2	US-10-104-047-2454	Sequence 2454, Ap
703	32	37.6	206	2	US-09-248-796A-26145	Sequence 26145, A	776	32	37.6	404	2	US-09-248-796A-15586	Sequence 15586, A
704	32	37.6	209	2	US-09-248-796A-15963	Sequence 15963, A	777	32	37.6	412	2	US-09-543-681A-8318	Sequence 8318, Ap
705	32	37.6	210	2	US-09-638-092-121	Sequence 121, App	778	32	37.6	420	2	US-09-844-006A-2	Sequence 2, Appli
706	32	37.6	210	2	US-09-487-558B-320	Sequence 320, App	779	32	37.6	427	2	US-08-506-296B-56	Sequence 56, Appli
707	32	37.6	211	2	US-09-323-872A-12	Sequence 12, Appli	780	32	37.6	431	2	US-08-807-342B-5	Sequence 5, Appli
708	32	37.6	211	2	US-09-072-433-21	Sequence 21, Appli	781	32	37.6	432	1	US-08-896-005-4	Sequence 4, Appli
709	32	37.6	213	2	US-09-302-540-10255	Sequence 10255, A	782	32	37.6	432	1	US-08-896-005-5	Sequence 5, Appli
710	32	37.6	220	2	US-09-270-767-42243	Sequence 42243, A	783	32	37.6	432	1	US-08-933-750C-47	Sequence 47, Appli
711	32	37.6	220	2	US-09-270-767-7522	Sequence 7522, A	784	32	37.6	432	2	US-09-234-613-47	Sequence 47, Appli
712	32	37.6	221	2	US-09-605-703B-2640	Sequence 2640, Ap	785	32	37.6	432	2	US-09-347-878-1	Sequence 1, Appli
713	32	37.6	231	2	US-09-248-796A-25146	Sequence 25146, A	786	32	37.6	432	2	US-09-546-013-3	Sequence 3, Appli
714	32	37.6	232	2	US-09-252-991A-27530	Sequence 27530, A	787	32	37.6	436	2	US-09-248-796A-20561	Sequence 20561, A
715	32	37.6	233	2	US-09-302-540-12466	Sequence 12466, A	788	32	37.6	438	2	US-09-457-046B-58	Sequence 58, Appli
716	32	37.6	234	2	US-09-270-767-42602	Sequence 42602, A	789	32	37.6	438	2	US-09-866-570B-58	Sequence 58, Appli
717	32	37.6	238	2	US-09-111-470-8	Sequence 8, Appli	790	32	37.6	439	2	US-09-792-420-7	Sequence 7, Appli
718	32	37.6	238	2	US-09-198-452A-727	Sequence 727, App	791	32	37.6	439	2	US-09-107-433-4731	Sequence 4731, Ap
719	32	37.6	238	2	US-09-862-802A-8	Sequence 8, Appli	792	32	37.6	441	2	US-09-107-532A-6658	Sequence 6658, Ap
720	32	37.6	238	2	US-09-438-185A-689	Sequence 689, App	793	32	37.6	441	2	US-09-792-420-6	Sequence 6, Appli
721	32	37.6	249	1	US-08-089-998B-2	Sequence 2, Appli	794	32	37.6	442	2	US-08-506-296B-70	Sequence 70, Appli
722	32	37.6	249	1	US-08-457-272-2	Sequence 2, Appli	795	32	37.6	442	2	US-09-710-279-1000	Sequence 1000, Ap
723	32	37.6	249	4	US-09-252-991A-26271	Sequence 2, Appli	796	32	37.6	443	2	US-09-457-046B-50	Sequence 50, Appli
724	32	37.6	249	4	PCT-US94-07595-2	Sequence 2, Appli	797	32	37.6	443	2	US-09-866-570B-50	Sequence 50, Appli
725	32	37.6	258	2	US-08-931-855B-6	Sequence 6, Appli	798	32	37.6	444	2	US-09-134-000C-3835	Sequence 3835, Ap
726	32	37.6	266	2	US-09-252-991A-24870	Sequence 24870, A	799	32	37.6	444	2	US-09-792-420-5	Sequence 5, Appli
727	32	37.6	269	2	US-09-640-211A-797	Sequence 797, App	800	32	37.6	447	2	US-09-248-796A-20541	Sequence 20541, A
728	32	37.6	274	1	US-08-127-499A-17	Sequence 17, Appli	801	32	37.6	447	2	US-09-949-016-10359	Sequence 10359, A
729	32	37.6	274	1	US-08-482-847-17	Sequence 17, Appli	802	32	37.6	456	2	US-09-489-039A-10363	Sequence 10363, A
730	32	37.6	275	2	US-09-051-755-4	Sequence 4, Appli	803	32	37.6	460	2	US-09-902-540-15589	Sequence 15589, A
731	32	37.6	275	2	US-09-302-540-15009	Sequence 15009, A	804	32	37.6	465	2	US-09-252-991A-29387	Sequence 29387, A
732	32	37.6	279	2	US-09-489-039A-9257	Sequence 9257, Ap	805	32	37.6	471	2	US-09-252-991A-20999	Sequence 20999, A
733	32	37.6	279	2	US-09-248-796A-17575	Sequence 17575, A	806	32	37.6	475	2	US-09-583-110-4327	Sequence 4327, Ap
734	32	37.6	280	1	US-08-781-562-4	Sequence 4, Appli	807	32	37.6	475	2	US-09-769-787-7	Sequence 7, Appli
735	32	37.6	287	2	US-09-107-532A-4780	Sequence 4780, Ap	808	32	37.6	476	2	US-09-107-433-3078	Sequence 3078, Ap
736	32	37.6	291	2	US-09-134-000C-3836	Sequence 3836, Ap	809	32	37.6	477	2	US-09-118-324-2	Sequence 2, Appli
737	32	37.6	297	2	US-09-107-532A-4719	Sequence 4719, Ap	810	32	37.6	483	2	US-09-538-092-1015	Sequence 1015, Ap
738	32	37.6	298	2	US-09-424-349A-7	Sequence 7, Appli	811	32	37.6	494	2	US-09-134-001C-4176	Sequence 4176, Ap
739	32	37.6	301	2	US-09-198-452A-1099	Sequence 1099, Ap	812	32	37.6	498	2	US-09-902-540-12285	Sequence 12285, A
740	32	37.6	302	2	US-09-457-046B-10	Sequence 10, Appli	813	32	37.6	500	1	US-08-896-005-1	Sequence 1, Appli
741	32	37.6	303	2	US-09-866-570B-10	Sequence 10, Appli	814	32	37.6	507	2	US-10-104-047-2009	Sequence 2009, Ap
742	32	37.6	303	2	US-09-198-452A-1084	Sequence 1084, Ap	815	32	37.6	512	2	US-09-248-796A-15186	Sequence 15186, A
743	32	37.6	304	2	US-08-506-296B-57	Sequence 57, Appli	816	32	37.6	513	2	US-09-303-518D-576	Sequence 576, App
744	32	37.6	311	2	US-09-543-681A-5602	Sequence 5602, Ap	817	32	37.6	513	2	US-09-303-518D-580	Sequence 580, App
745	32	37.6	323	2	US-09-252-991A-24673	Sequence 24673, A	818	32	37.6	513	2	US-09-303-518D-582	Sequence 582, App
746	32	37.6	328	2	US-09-438-185A-1025	Sequence 1025, Ap	819	32	37.6	527	2	US-09-248-796A-15395	Sequence 15395, A
747	32	37.6	331	1	US-08-907-674-1	Sequence 1, Appli	820	32	37.6	529	2	US-10-104-047-2137	Sequence 2137, Ap
748	32	37.6	331	1	US-09-215-087-1	Sequence 1, Appli	821	32	37.6	542	2	US-08-506-296B-69	Sequence 69, Appli
749	32	37.6	331	2	US-09-391-959-1	Sequence 1, Appli	822	32	37.6	542	2	US-09-389-956-78	Sequence 78, Appli
750	32	37.6	332	2	US-09-537-654-8	Sequence 8, Appli	823	32	37.6	547	2	US-09-605-703B-2398	Sequence 2398, Ap
751	32	37.6	332	2	US-09-602-777A-308	Sequence 308, App	824	32	37.6	553	2	US-09-270-767-43146	Sequence 43146, A
752	32	37.6	334	2	US-09-489-039A-11523	Sequence 11523, A	825	32	37.6	555	2	US-09-999-833A-109	Sequence 109, App
753	32	37.6	336	2	US-09-095-163-2	Sequence 2, Appli	826	32	37.6	555	2	US-10-020-445A-109	Sequence 109, App
754	32	37.6	338	2	US-09-107-532A-5819	Sequence 5819, Ap	827	32	37.6	562	2	US-09-487-558B-82	Sequence 82, Appli
755	32	37.6	339	2	US-09-248-796A-19599	Sequence 19599, A	828	32	37.6	562	6	5258502-2	Patent No. 5258502
756	32	37.6	342	2	US-09-502-540-15143	Sequence 15143, A	829	32	37.6	583	2	Sequence 8267, Ap	Sequence 8267, Ap
757	32	37.6	345	2	US-09-328-352-7063	Sequence 7063, Ap	830	32	37.6	586	2	US-10-104-047-3485	Sequence 3485, Ap



831	32	37.6	589	2	US-09-377-155-19	Sequence 19, Appl	904	32	37.6	1755	2	US-09-724-126A-4	Sequence 4, Appl
832	32	37.6	589	2	US-09-669-374-19	Sequence 19, Appl	905	32	37.6	1755	2	US-09-724-126A-6	Sequence 6, Appl
833	32	37.6	589	2	US-09-797-862-19	Sequence 19, Appl	906	32	37.6	1755	2	US-09-693-205A-8	Sequence 8, Appl
834	32	37.6	589	2	US-09-302-626B-108	Sequence 108, App	907	32	37.6	2332	1	US-07-864-004B-4	Sequence 4, Appl
835	32	37.6	589	2	US-09-302-626B-109	Sequence 109, App	908	32	37.6	2332	1	US-08-251-937A-4	Sequence 4, Appl
836	32	37.6	595	2	US-09-302-626B-97	Sequence 97, Appl	909	32	37.6	2332	1	US-08-212-133A-2	Sequence 2, Appl
837	32	37.6	595	2	US-09-248-796A-14562	Sequence 14562, A	910	32	37.6	2332	1	US-08-276-594A-2	Sequence 2, Appl
838	32	37.6	598	2	US-09-650-324A-58	Sequence 58, Appl	911	32	37.6	2332	1	US-08-474-503-2	Sequence 2, Appl
839	32	37.6	598	2	US-09-613-486-36	Sequence 36, Appl	912	32	37.6	2332	1	US-08-670-707A-2	Sequence 2, Appl
840	32	37.6	598	2	US-10-039-112A-58	Sequence 58, Appl	913	32	37.6	2332	1	US-09-037-601-2	Sequence 2, Appl
841	32	37.6	600	2	US-09-438-185A-1013	Sequence 1013, Ap	914	32	37.6	2332	2	US-09-324-867-3	Sequence 3, Appl
842	32	37.6	608	2	US-10-003-392-2	Sequence 2, Appl	915	32	37.6	2332	2	US-09-315-179-2	Sequence 2, Appl
843	32	37.6	616	1	US-08-749-882A-2	Sequence 2, Appl	916	32	37.6	2332	2	US-09-523-656-2	Sequence 2, Appl
844	32	37.6	616	1	US-08-539-134-2	Sequence 2, Appl	917	32	37.6	2332	2	US-09-957-641A-2	Sequence 2, Appl
845	32	37.6	616	1	US-08-991-531-2	Sequence 2, Appl	918	32	37.6	2332	2	US-10-360-101-229	Sequence 229, App
846	32	37.6	616	1	US-09-032-315-9	Sequence 9, Appl	919	32	37.6	2332	4	PCT-US93-03275-A	Sequence 4, Appl
847	32	37.6	616	1	US-08-993-318A-9	Sequence 9, Appl	920	32	37.6	2332	4	PCT-US94-13200-2	Sequence 2, Appl
848	32	37.6	616	2	US-09-028-887-2	Sequence 2, Appl	921	32	37.6	2351	1	US-08-121-202-2	Sequence 2, Appl
849	32	37.6	616	2	US-09-399-886-9	Sequence 9, Appl	922	32	37.6	2351	1	US-08-366-851A-2	Sequence 2, Appl
850	32	37.6	616	2	US-09-396-260-9	Sequence 9, Appl	923	32	37.6	2351	2	US-10-133-907-4	Sequence 4, Appl
851	32	37.6	616	2	US-09-518-901-2	Sequence 2, Appl	924	32	37.6	2351	2	US-09-001-039B-45	Sequence 45, Appl
852	32	37.6	616	2	US-09-576-281-9	Sequence 9, Appl	925	32	37.6	2351	6	5171844-2	Patent No. 5171844
853	32	37.6	616	4	PCT-US95-05815-2	Sequence 5348, Ap	926	32	37.6	2351	6	5422260-1	Patent No. 5422260
854	32	37.6	616	2	US-09-328-352-5348	Sequence 5348, Ap	927	32	37.6	3079	4	PCT-US94-00198-4	Sequence 4, Appl
855	32	37.6	621	2	US-09-026-343-7	Sequence 7, Appl	928	32	37.6	3135	1	US-08-323-170B-2	Sequence 2, Appl
856	32	37.6	621	2	US-09-362-871-7	Sequence 7, Appl	929	32	37.6	3135	2	US-08-954-441-2	Sequence 2, Appl
857	32	37.6	632	2	US-09-252-991A-23096	Sequence 23096, A	930	32	37.6	3829	2	US-09-693-205A-2	Sequence 2, Appl
858	32	37.6	638	2	US-09-134-006A-4864	Sequence 4864, Ap	931	32	37.6	3829	2	US-09-693-205A-16	Sequence 16, Appl
859	32	37.6	645	2	US-09-248-796A-20288	Sequence 20288, A	932	32	37.6	3830	2	US-09-693-205A-4	Sequence 4, Appl
860	32	37.6	663	2	US-10-104-047-3473	Sequence 3473, Ap	933	31.5	37.1	51	1	US-08-870-518-20	Sequence 20, Appl
861	32	37.6	665	2	US-08-506-296B-68	Sequence 68, Appl	934	31.5	37.1	85	2	US-08-858-207A-474	Sequence 474, App
862	32	37.6	665	2	US-09-769-787-127	Sequence 127, App	935	31.5	37.1	132	2	US-08-754-477A-135	Sequence 135, App
863	32	37.6	668	2	US-09-248-796A-17495	Sequence 17495, A	936	31.5	37.1	143	2	US-09-414-276-4	Sequence 4, Appl
864	32	37.6	677	2	US-09-252-991A-32924	Sequence 32924, A	937	31.5	37.1	195	2	US-09-248-796A-20608	Sequence 20608, A
865	32	37.6	679	2	US-09-949-016-8034	Sequence 8034, Ap	938	31.5	37.1	224	2	US-09-248-796A-17768	Sequence 17768, A
866	32	37.6	679	2	US-09-949-016-8035	Sequence 8035, Ap	939	31.5	37.1	234	2	US-09-543-681A-7778	Sequence 7778, Ap
867	32	37.6	680	2	US-09-134-000C-4799	Sequence 4799, Ap	940	31.5	37.1	255	2	US-09-949-016-7430	Sequence 7430, Ap
868	32	37.6	694	2	US-09-248-796A-14448	Sequence 14448, A	941	31.5	37.1	269	2	US-09-902-540-14940	Sequence 14940, A
869	32	37.6	701	2	US-09-538-092-303	Sequence 303, App	942	31.5	37.1	271	2	US-08-957-351-26	Sequence 26, Appl
870	32	37.6	720	2	US-09-252-991A-31915	Sequence 31915, A	943	31.5	37.1	271	2	US-08-754-477A-2	Sequence 2, Appl
871	32	37.6	728	2	US-09-252-991A-23738	Sequence 23738, A	944	31.5	37.1	271	2	US-08-754-477A-5	Sequence 5, Appl
872	32	37.6	748	2	US-09-252-991A-32942	Sequence 32942, A	945	31.5	37.1	271	2	US-08-754-477A-121	Sequence 121, App
873	32	37.6	761	2	US-09-328-352-5942	Sequence 5942, Ap	946	31.5	37.1	271	2	US-08-754-477A-135	Sequence 135, App
874	32	37.6	806	1	US-07-980-528-2	Sequence 2, Appl	947	31.5	37.1	271	2	US-08-754-477A-131	Sequence 131, App
875	32	37.6	824	2	US-09-538-092-1242	Sequence 1242, Ap	948	31.5	37.1	271	2	US-09-949-016-5936	Sequence 5936, Ap
876	32	37.6	836	2	US-09-489-039A-7701	Sequence 7701, Ap	949	31.5	37.1	277	2	US-09-248-796A-15054	Sequence 15054, A
877	32	37.6	848	2	US-09-252-991A-27842	Sequence 27842, A	950	31.5	37.1	293	1	US-08-701-191A-16	Sequence 16, Appl
878	32	37.6	877	2	US-08-917-441-102	Sequence 102, App	951	31.5	37.1	293	2	US-09-664-526-16	Sequence 16, Appl
879	32	37.6	878	2	US-09-463-238-4	Sequence 4, Appl	952	31.5	37.1	305	2	US-08-558-135-6	Sequence 6, Appl
880	32	37.6	883	2	US-09-463-238-19	Sequence 19, Appl	953	31.5	37.1	317	2	US-09-540-236-3830	Sequence 3830, Ap
881	32	37.6	898	2	US-09-302-540-11561	Sequence 11561, A	954	31.5	37.1	333	2	US-09-134-000C-5488	Sequence 5488, Ap
882	32	37.6	916	2	US-09-248-796A-25756	Sequence 25756, A	955	31.5	37.1	335	2	US-09-414-276-6	Sequence 6, Appl
883	32	37.6	1040	2	US-09-564-805-238	Sequence 238, App	956	31.5	37.1	386	2	US-09-390-326-6	Sequence 6, Appl
884	32	37.6	1123	2	US-09-408-865-1	Sequence 1, Appl	957	31.5	37.1	386	2	US-09-939-833-6	Sequence 6, Appl
885	32	37.6	1125	2	US-09-500-920-60	Sequence 60, Appl	958	31.5	37.1	386	2	US-09-506-906-6	Sequence 6, Appl
886	32	37.6	1148	2	US-09-538-092-156	Sequence 156, App	959	31.5	37.1	386	2	US-09-939-832-6	Sequence 6, Appl
887	32	37.6	1160	2	US-09-711-164-401	Sequence 401, App	960	31.5	37.1	386	2	US-09-939-754-6	Sequence 6, Appl
888	32	37.6	1161	2	US-09-252-991A-29048	Sequence 29048, A	961	31.5	37.1	425	2	US-09-248-796A-20774	Sequence 20774, A
889	32	37.6	1178	2	US-09-489-039A-9944	Sequence 9944, Ap	962	31.5	37.1	431	2	US-09-586-305A-18	Sequence 18, Appl
890	32	37.6	1189	2	US-09-489-039A-13776	Sequence 13776, A	963	31.5	37.1	435	2	US-09-162-021B-12	Sequence 12, Appl
891	32	37.6	1202	2	US-09-328-352-6889	Sequence 6889, Ap	964	31.5	37.1	438	2	US-09-586-305A-11	Sequence 11, Appl
892	32	37.6	1203	2	US-09-489-039A-12326	Sequence 12326, A	965	31.5	37.1	438	2	US-09-586-305A-12	Sequence 12, Appl
893	32	37.6	1240	2	US-09-538-092-658	Sequence 658, App	966	31.5	37.1	438	2	US-09-586-305A-13	Sequence 13, Appl
894	32	37.6	1253	2	US-08-506-296B-14	Sequence 14, Appl	967	31.5	37.1	438	2	US-09-586-305A-14	Sequence 14, Appl
895	32	37.6	1295	2	US-09-248-796A-18447	Sequence 18447, A	968	31.5	37.1	438	2	US-09-586-305A-15	Sequence 15, Appl
896	32	37.6	1438	2	US-09-509-916-1	Sequence 1, Appl	969	31.5	37.1	438	2	US-09-586-305A-16	Sequence 16, Appl
897	32	37.6	1445	2	US-09-407-605-3	Sequence 3, Appl	970	31.5	37.1	438	2	US-09-586-305A-17	Sequence 17, Appl
898	32	37.6	1447	2	US-09-407-605-4	Sequence 4, Appl	971	31.5	37.1	438	2	US-09-586-305A-19	Sequence 19, Appl
899	32	37.6	1457	2	US-09-001-039B-47	Sequence 47, Appl	972	31.5	37.1	438	2	US-09-586-305A-20	Sequence 20, Appl
900	32	37.6	1471	1	US-08-683-839B-3	Sequence 3, Appl	973	31.5	37.1	447	2	US-09-070-356-7	Sequence 7, Appl
901	32	37.6	1661	1	US-08-982-083-2	Sequence 2, Appl	974	31.5	37.1	496	2	US-09-292-768-66	Sequence 66, Appl
902	32	37.6	1661	1	US-08-558-107-2	Sequence 2, Appl	975	31.5	37.1	496	2	US-09-270-767-42845	Sequence 42845, A
903	32	37.6	1661	2	US-09-243-539-2	Sequence 2, Appl	976	31.5	37.1	570	2	US-08-747-221B-54	Sequence 54, Appl

977 31.5 37.1 570 2 US-08-747-221B-55 Sequence 55, Appl  
978 31.5 37.1 570 2 US-09-005-051-54 Sequence 54, Appl  
979 31.5 37.1 570 2 US-09-005-051-55 Sequence 55, Appl  
980 31.5 37.1 570 2 US-09-403-942F-54 Sequence 54, Appl  
981 31.5 37.1 570 2 US-09-403-942F-55 Sequence 55, Appl  
982 31.5 37.1 595 2 US-08-747-221B-31 Sequence 31, Appl  
983 31.5 37.1 595 2 US-09-005-051-31 Sequence 31, Appl  
984 31.5 37.1 595 2 US-09-403-942F-31 Sequence 31, Appl  
985 31.5 37.1 596 2 US-08-747-221B-25 Sequence 25, Appl  
986 31.5 37.1 596 2 US-09-005-051-25 Sequence 25, Appl  
987 31.5 37.1 596 2 US-09-403-942F-25 Sequence 25, Appl  
988 31.5 37.1 880 2 US-09-538-092-441 Sequence 441, App  
989 31.5 37.1 880 2 US-09-487-558B-440 Sequence 440, App  
990 31.5 37.1 1003 1 US-08-571-758-4 Sequence 4, Appl  
991 31.5 37.1 1003 1 US-08-909-984A-4 Sequence 4, Appl  
992 31.5 37.1 1003 1 US-08-909-983-4 Sequence 4, Appl  
993 31.5 37.1 1088 2 US-09-961-403-4 Sequence 4, Appl  
994 31.5 37.1 1089 1 US-08-180-195-36 Sequence 36, Appl  
995 31.5 37.1 1089 1 US-08-168-917-4 Sequence 4, Appl  
996 31.5 37.1 1089 1 US-08-777-329-36 Sequence 36, Appl  
997 31.5 37.1 1089 1 US-08-475-458-36 Sequence 36, Appl  
998 31.5 37.1 1089 1 US-08-460-510-4 Sequence 4, Appl  
999 31.5 37.1 1089 1 US-08-460-490-4 Sequence 4, Appl  
1000 31.5 37.1 1089 2 US-08-980-400-36 Sequence 36, Appl

## ALIGNMENTS

RESULT 1  
US-08-057-162B-5  
; Sequence 5, Application US/08057162B  
; Patent No. 5686075  
; GENERAL INFORMATION:  
; APPLICANT: Taubman, Martin A.  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR DENTAL CARIES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/057,162B  
; FILING DATE: 30-APR-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/877,295  
; FILING DATE: 01-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wagner, Richard W.  
; REGISTRATION NUMBER: 34,480  
; REFERENCE/DOCKET NUMBER: FDC92-01A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 14 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-057-162B-5

Query Match 100.0%; Score 85; DB 1; Length 14;  
Best Local Similarity 100.0%; Pred. No. 6.3e-08;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDHL 14  
Db 1 QWNGESEKPYDDHL 14  
|||||

RESULT 2  
US-09-290-049A-4  
; Sequence 4, Application US/092900049A  
; Patent No. 6827936  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: MAC peptide  
US-09-290-049A-4

Query Match 100.0%; Score 85; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 6.8e-08;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDHL 14  
Db 2 QWNGESEKPYDDHL 15  
|||||

RESULT 3  
US-09-290-049A-8  
; Sequence 8, Application US/092900049A  
; Patent No. 6827936  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: S. downei  
US-09-290-049A-8

Query Match 100.0%; Score 85; DB 2; Length 15;  
Best Local Similarity 100.0%; Pred. No. 6.8e-08;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWNGESEKPYDDHL 14

```
Db      2 QWNGSEKPYDDHL 15
|||||
RESULT 4
US-09-290-049A-9
; Sequence 9, Application US/09290049A
; Patent No. 6827936
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 15
; TYPE: PRT
; ORGANISM: S. sobrinus
US-09-290-049A-9

Query Match      100.0%; Score 85; DB 2; Length 15;
Best Local Similarity 100.0%; Pred. No. 6.8e-08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QWNGSEKPYDDHL 14
|||||
Db      2 QWNGSEKPYDDHL 15
|||||
RESULT 5
US-09-290-049A-6
; Sequence 6, Application US/09290049A
; Patent No. 6827936
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; TITLE OF INVENTION: CRIES
; FILE REFERENCE: 1564.1008-002
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 15
; TYPE: PRT
; ORGANISM: S. mutans
US-09-290-049A-6

Query Match      78.8%; Score 67; DB 2; Length 15;
Best Local Similarity 76.9%; Pred. No. 7.6e-05;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      2 QWNGSEKPYDDHL 14
|||||
Db      3 WNSDSEKPFDDHL 15
|||||
RESULT 6
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match      78.8%; Score 67; DB 2; Length 1375;
Best Local Similarity 76.9%; Pred. No. 0.012;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy      2 WNGSEKPYDDHL 14
|||||
Db      370 WNSDSEKPFDDHL 382
|||||
RESULT 7
US-09-740-274-4
; Sequence 4, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-4

Query Match      78.8%; Score 67; DB 2; Length 1375;
Best Local Similarity 76.9%; Pred. No. 0.012;
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

QY 2 WNGSEKPYDDHL 14  
|| :|||:||||  
Db 370 WNSDSEKPFDDHL 382

RESULT 8  
US-09-007-999-2  
; Sequence 2, Application US/09007999  
; Patent No. 6087559  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0356D  
; CURRENT APPLICATION NUMBER: US/09/007,999  
; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-007-999-2

Query Match 78.8%; Score 67; DB 2; Length 1475;  
Best Local Similarity 76.9%; Pred. No. 0.013;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14  
|| :|||:||||  
Db 344 WNSDSEKPFDDHL 356

RESULT 9  
US-09-210-361-2  
; Sequence 2, Application US/09210361  
; Patent No. 6284479  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starches and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0357CR  
; CURRENT APPLICATION NUMBER: US/09/210,361  
; CURRENT FILING DATE: 1998-12-11  
; EARLIER APPLICATION NUMBER: 09/007,999  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/478,704  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/009,620  
; EARLIER FILING DATE: 1998-01-20  
; EARLIER APPLICATION NUMBER: 08/485,243  
; EARLIER FILING DATE: 1995-06-07  
; EARLIER APPLICATION NUMBER: 09/008,172  
; EARLIER FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-210-361-2

Query Match 78.8%; Score 67; DB 2; Length 1475;  
Best Local Similarity 76.9%; Pred. No. 0.013;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14  
|| :|||:||||  
Db 344 WNSDSEKPFDDHL 356

RESULT 10  
US-09-740-274-2  
; Sequence 2, Application US/09740274  
; Patent No. 6465203  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSEQ for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 78.8%; Score 67; DB 2; Length 1475;  
Best Local Similarity 76.9%; Pred. No. 0.013;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14  
|| :|||:||||  
Db 344 WNSDSEKPFDDHL 356

RESULT 11  
US-09-290-049A-5  
; Sequence 5, Application US/09290049A  
; Patent No. 6827936  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL  
; TITLE OF INVENTION: CARIES  
; FILE REFERENCE: 1564.1008-002  
; CURRENT APPLICATION NUMBER: US/09/290,049A  
; CURRENT FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: S. mutans  
US-09-290-049A-5

Query Match 75.3%; Score 64; DB 2; Length 15;  
Best Local Similarity 69.2%; Pred. No. 0.00024;  
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14  
|| :|||:||||

```
Db          3 WNSDSRPFDDHL 15

RESULT 12
US-09-543-681A-4957
; Sequence 4957, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4957
; LENGTH: 357
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4957

Query Match          55.9%; Score 47.5; DB 2; Length 357;
Best Local Similarity 47.4%; Pred. No. 5.3;
Matches 9; Conservative 1; Mismatches 2; Indels 7; Gaps 1;

QY      2 WNGSEKPP-----YDDH 13
      |||: |||
Db      40 WNGKDEKQIATLPSYDPH 58

RESULT 13
US-09-543-681A-8307
; Sequence 8307, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8307
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8307

Query Match          55.3%; Score 47; DB 2; Length 106;
Best Local Similarity 53.8%; Pred. No. 1.6;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      1 QWNGSEKPYDDH 13
      :|||: |||
Db      4 EWNGEYISPYAEH 16

RESULT 14
US-09-134-001C-2980
; Sequence 2980, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2980
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2980

Query Match          54.1%; Score 46; DB 2; Length 504;
Best Local Similarity 54.5%; Pred. No. 14;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      2 WNGSEKPYDD 12
      |||: |||
Db      420 WRGTNDKPYED 430

RESULT 15
US-09-489-039A-12202
; Sequence 12202, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 12202
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12202

Query Match          49.4%; Score 42; DB 2; Length 187;
Best Local Similarity 46.2%; Pred. No. 22;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      1 QWNGSEKPYDDH 13
      :|||: |||
Db      85 EWSGEYISPYAEH 97

RESULT 16
US-09-252-991A-30733
; Sequence 30733, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30733
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30733

Query Match          49.4%; Score 42; DB 2; Length 271;
Best Local Similarity 38.5%; Pred. No. 33;
Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;
```

QY 2 WNGSEKPYDDHL 14  
Db 195 WNGDPORPEARL 207

RESULT 17  
US-09-583-110-3119  
; Sequence 3119, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 3119  
; LENGTH: 424  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-3119

Query Match 49.4%; Score 42; DB 2; Length 424;  
Best Local Similarity 53.8%; Pred. No. 54;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14  
Db 5 WNGFSKKSQERL 17

RESULT 18  
US-09-107-433-3690  
; Sequence 3690, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,433  
; FILING DATE: 30-Jun-1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/ 085131  
; FILING DATE: May 12, 1998  
; APPLICATION NUMBER: 60/051553  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-011

; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 3690:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 426 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (B) LOCATION 1...426  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3690:  
US-09-107-433-3690

Query Match 49.4%; Score 42; DB 2; Length 426;  
Best Local Similarity 53.8%; Pred. No. 55;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14  
Db 7 WNGFSKKSQERL 19

RESULT 19  
US-09-543-681A-4480  
; Sequence 4480, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 4480  
; LENGTH: 473  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-4480

Query Match 48.2%; Score 41; DB 2; Length 473;  
Best Local Similarity 75.0%; Pred. No. 91;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QWGESEK 8  
Db 451 QWNGQTEK 458

RESULT 20  
US-09-270-767-34087  
; Sequence 34087, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 34087  
; LENGTH: 189  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-34087

```
Query Match      47.1%; Score 40; DB 2; Length 189;
Best Local Similarity 87.5%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 NGESEKPY 10
      ||||| ||
Db      137 NGESESPY 144

RESULT 21
US-09-270-767-49304
; Sequence 49304, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 49304
; LENGTH: 189
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-49304

Query Match      47.1%; Score 40; DB 2; Length 189;
Best Local Similarity 87.5%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      3 NGESEKPY 10
      ||||| ||
Db      137 NGESESPY 144

RESULT 22
US-09-134-000C-3729
; Sequence 3729, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6912
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3729
; LENGTH: 252
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-3729

Query Match      47.1%; Score 40; DB 2; Length 252;
Best Local Similarity 66.7%; Pred. No. 66;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 QWNGSEKPY 9
      ||| |::|||
Db      238 QWNGQTEKP 246

RESULT 23
US-09-710-279-2760
; Sequence 2760, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
```

```
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU34800S
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2760
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-2760

Query Match      47.1%; Score 40; DB 2; Length 319;
Best Local Similarity 63.6%; Pred. No. 86;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      2 WNGESEKPYDD 12
      :||| |||:
Db      46 FNKESEDPYDE 56

RESULT 24
US-09-134-001C-3183
; Sequence 3183, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3183
; LENGTH: 443
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3183

Query Match      47.1%; Score 40; DB 2; Length 443;
Best Local Similarity 63.6%; Pred. No. 1.2e+02;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      2 WNGESEKPYDD 12
      :||| |||:
Db      170 FNKESEDPYDE 180

RESULT 25
US-09-487-558B-360
; Sequence 360, Application US/09487558B
; Patent No. 6949356
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. 6949356man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
```

```
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.130
; CURRENT APPLICATION NUMBER: US/09/487,558B
; CURRENT FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/487,558
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 446
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 360
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Penicillium chrysogenum
US-09-487-558B-360

Query Match          47.1%; Score 40; DB 2; Length 532;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      1 OWNGESEKP 9
       | | | | | : | |
Db      67 QWNGHNETP 75

RESULT 26
US-09-270-767-33478
; Sequence 33478, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33478
; LENGTH: 107
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-33478

Query Match          45.9%; Score 39; DB 2; Length 107;
Best Local Similarity 60.0%; Pred. No. 37;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      3 NGESEKPYDD 12
       | : | : | | | |
Db      97 NPKTDKPYDD 106

RESULT 27
US-09-710-279-356
; Sequence 356, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 356
; LENGTH: 309
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
```

```
US-09-710-279-356

Query Match          45.9%; Score 39; DB 2; Length 309;
Best Local Similarity 54.5%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      2 WNGESEKPYDD 12
       : | | | : | : |
Db      99 YRGESDLPFDD 109

RESULT 28
US-09-134-001C-4371
; Sequence 4371, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4371
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4371

Query Match          45.9%; Score 39; DB 2; Length 313;
Best Local Similarity 54.5%; Pred. No. 1.2e+02;
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      2 WNGESEKPYDD 12
       : | | | : | : |
Db      103 YRGESDLPFDD 113

RESULT 29
US-09-117-415B-2
; Sequence 2, Application US/09117415B
; Patent No. 6551586
; GENERAL INFORMATION:
; APPLICANT: Davidson, Eugene
; TITLE OF INVENTION: Malaria Vaccine Based Upon the Addition
; OF a MSA1 Peptide
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henry D. Coleman, COLEMAN SUDOL SAPONE, PC
; STREET: 714 Colorado Avenue
; CITY: Bridgeport
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPad (ASCII)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/117,415B
; FILING DATE: 29-Jul-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coleman, Henry D.
; REGISTRATION NUMBER: 32,559
; REFERENCE/DOCKET NUMBER: R12-030
; TELECOMMUNICATION INFORMATION:
```





```
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 4 GESEKPYDD 12
Db 257 GESEEDYDD 265

RESULT 33
US-09-117-415B-20
; Sequence 20, Application US/09117415B
; Patent No. 6551586
; GENERAL INFORMATION:
; APPLICANT: Davidson, Eugene
; Yang, Shutong
; TITLE OF INVENTION: Malaria Vaccine Based Upon the Addition
; of a MSA1 Peptide
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henry D. Coleman, COLEMAN SUDOL SAPONE, PC
; STREET: 714 Colorado Avenue
; CITY: Bridgeport
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPad (ASCII)
; CURRENT APPLICATION DATA:
; FILING DATE: 29-Jul-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coleman, Henry D.
; REGISTRATION NUMBER: 32,559
; REFERENCE/DOCKET NUMBER: R12-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 679-0090
; TELEFAX: (212) 679-9121
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 631 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-117-415B-20
Query Match 45.9%; Score 39; DB 2; Length 631;
Best Local Similarity 77.8%; Pred. No. 2.7e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 4 GESEKPYDD 12
Db 239 GESEEDYDD 247

RESULT 34
US-09-117-415B-16
; Sequence 16, Application US/09117415B
; Patent No. 6551586
; GENERAL INFORMATION:
; APPLICANT: Davidson, Eugene
; Yang, Shutong
; TITLE OF INVENTION: Malaria Vaccine Based Upon the Addition
; of a MSA1 Peptide
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henry D. Coleman, COLEMAN SUDOL SAPONE, PC
; STREET: 714 Colorado Avenue
; CITY: Bridgeport
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPad (ASCII)
; CURRENT APPLICATION DATA:
; FILING DATE: 29-Jul-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coleman, Henry D.
; REGISTRATION NUMBER: 32,559
; REFERENCE/DOCKET NUMBER: R12-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 679-0090
; TELEFAX: (212) 679-9121
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 631 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-117-415B-20
Query Match 45.9%; Score 39; DB 2; Length 631;
Best Local Similarity 77.8%; Pred. No. 2.7e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 4 GESEKPYDD 12
Db 239 GESEEDYDD 247

RESULT 34
US-09-117-415B-16
; Sequence 16, Application US/09117415B
; Patent No. 6551586
; GENERAL INFORMATION:
; APPLICANT: Davidson, Eugene
; Yang, Shutong
; TITLE OF INVENTION: Malaria Vaccine Based Upon the Addition
; of a MSA1 Peptide
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Henry D. Coleman, COLEMAN SUDOL SAPONE, PC
; STREET: 714 Colorado Avenue
; CITY: Bridgeport
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 10017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPad (ASCII)
; CURRENT APPLICATION DATA:
; FILING DATE: 29-Jul-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coleman, Henry D.
; REGISTRATION NUMBER: 32,559
; REFERENCE/DOCKET NUMBER: R12-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 679-0090
; TELEFAX: (212) 679-9121
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 631 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-09-117-415B-16
Query Match 45.9%; Score 39; DB 2; Length 649;
Best Local Similarity 77.8%; Pred. No. 2.8e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 4 GESEKPYDD 12
Db 257 GESEEDYDD 265

RESULT 35
US-09-269-874A-7
; Sequence 7, Application US/09269874A
; Patent No. 6933130
; GENERAL INFORMATION:
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Recombinant Process for Preparing a
; FILE OF INVENTION: Complete Malaria Antigen, GPI90/MSPI
; FILE REFERENCE: GRUE-003
; CURRENT APPLICATION NUMBER: US/09/269,874A
; CURRENT FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 1602
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-269-874A-7
Query Match 45.9%; Score 39; DB 2; Length 1602;
Best Local Similarity 77.8%; Pred. No. 7.8e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 4 GESEKPYDD 12
Db 1228 GESEEDYDD 1236

RESULT 36
US-09-269-874A-5
; Sequence 5, Application US/09269874A
; Patent No. 6933130
; GENERAL INFORMATION:
```

```
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Recombinant Process for Preparing a
; FILE REFERENCE: GRUE-003
; CURRENT APPLICATION NUMBER: US/09/269,874A
; CURRENT FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1621
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-269-874A-5

Query Match      45.9%; Score 39; DB 2; Length 1621;
Best Local Similarity 77.8%; Pred. No. 7.9e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      4 GESEKPYDD 12
Db      1247 GESEEDYDD 1255

RESULT 37
US-09-269-874A-3
; Sequence 3, Application US/09269874A
; Patent No. 6933130
; GENERAL INFORMATION:
; APPLICANT: Bujard, Hermann
; TITLE OF INVENTION: Recombinant Process for Preparing a
; FILE REFERENCE: GRUE-003
; CURRENT APPLICATION NUMBER: US/09/269,874A
; CURRENT FILING DATE: 1999-08-02
; PRIOR APPLICATION NUMBER: PCT/EP97/05441
; PRIOR FILING DATE: 1997-10-02
; PRIOR APPLICATION NUMBER: DE 19640817.2
; PRIOR FILING DATE: 1996-10-02
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1639
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
US-09-269-874A-3

Query Match      45.9%; Score 39; DB 2; Length 1639;
Best Local Similarity 77.8%; Pred. No. 8e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      4 GESEKPYDD 12
Db      1247 GESEEDYDD 1255

RESULT 38
US-09-134-000C-4510
; Sequence 4510, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
```

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4510
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-4510

Query Match      44.7%; Score 38; DB 2; Length 124;
Best Local Similarity 46.2%; Pred. No. 65;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY      1 QWNGESEKPYDDH 13
Db      94 EYKVESEEPYAEH 106

RESULT 39
US-09-540-236-2975
; Sequence 2975, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR:
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2975
; LENGTH: 144
; TYPE: PRT
; ORGANISM: M.catarrhalis
US-09-540-236-2975

Query Match      44.7%; Score 38; DB 2; Length 144;
Best Local Similarity 70.0%; Pred. No. 77;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY      5 ESEKPYDDHL 14
Db      93 ELEKPNDHL 102

RESULT 40
US-09-543-681A-7478
; Sequence 7478, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7478
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7478

Query Match      44.7%; Score 38; DB 2; Length 269;
Best Local Similarity 36.4%; Pred. No. 1.5e+02;
Matches 4; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY      1 QWNGESEKPYD 11
Db      183 QWKNDNDPFD 193

RESULT 41
```



Query Match 44.7%; Score 38; DB 2; Length 638;  
Best Local Similarity 41.7%; Pred. No. 4.1e+02;  
Matches 5; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 NGESEKPYDDHL 14  
Db 485 NGNDQYDSHI 496

RESULT 46  
US-08-851-567B-34  
; Sequence 34, Application US/08851567B  
; Patent No. 6528484  
; GENERAL INFORMATION:  
; APPLICANT: Ensign, Jerald C  
; APPLICANT: Bowen, David J  
; APPLICANT: Petell, James  
; APPLICANT: Fatig, Raymond  
; APPLICANT: Schoonover, Sue  
; APPLICANT: Iffrench-Constant, Richard  
; APPLICANT: Rocheleau, Thomas A.  
; APPLICANT: Blackburn, Michael B.  
; APPLICANT: Hey, Timothy D.  
; APPLICANT: Merlo, Donald J.  
; APPLICANT: Orr, Gregory L.  
; APPLICANT: Roberts, Jean L.  
; APPLICANT: Strickland, James A.  
; APPLICANT: Guo, Lining  
; APPLICANT: Ciche, Todd A.  
; APPLICANT: Sukhupinda, Kitieri  
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus  
; NUMBER OF SEQUENCES: 88  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dow AgroSciences Patent Department  
; STREET: 9330 Zionsville Road  
; CITY: Indianapolis  
; STATE: IN  
; COUNTRY: US  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/851,567B  
; FILING DATE: 05-MAY-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/063,615  
; FILING DATE: 18-MAY-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/395,497  
; FILING DATE: 28-FEB-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/007,255  
; FILING DATE: 06-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/608,423  
; FILING DATE: 28-FEB-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/705,484  
; FILING DATE: 28-AUG-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Seay, Nicholas J  
; REGISTRATION NUMBER: 27386  
; REFERENCE/DOCKET NUMBER: 960296.93804  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 608-251-5000  
; TELEFAX: 608-251-9166  
; INFORMATION FOR SEQ ID NO: 34:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1095 amino acids

; TYPE: amino acids  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-851-567B-34

Query Match 44.7%; Score 38; DB 2; Length 1095;  
Best Local Similarity 100.0%; Pred. No. 7.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 PYDDHL 14  
Db 205 PYDDHL 210

RESULT 47  
US-09-270-767-43051  
; Sequence 43051, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 43051  
; LENGTH: 349  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-43051

Query Match 44.1%; Score 37.5; DB 2; Length 349;  
Best Local Similarity 53.3%; Pred. No. 2.5e+02;  
Matches 8; Conservative 1; Mismatches 1; Indels 5; Gaps 1;

QY 3 NGESE-----KPYDD 12  
Db 173 NGEELSQYRPYDD 187

RESULT 48  
US-09-445-472-1  
; Sequence 1, Application US/09445472  
; Patent No. 6358726  
; GENERAL INFORMATION:  
; APPLICANT: TAKAKURA, Hikaru  
; APPLICANT: MORISHITA, Mio  
; APPLICANT: SHIMOJO, Tomoko  
; APPLICANT: ASADA, Kiyozo  
; APPLICANT: KATO, Ikunoshin  
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE  
; FILE REFERENCE: TAKAKURA=6  
; CURRENT APPLICATION NUMBER: US/09/445,472  
; CURRENT FILING DATE: 1999-12-06  
; PRIOR APPLICATION NUMBER: 151969/1997  
; PRIOR FILING DATE: 1997-06-10  
; NUMBER OF SEQ ID NOS: 33  
; SOFTWARE: Patent In version 3.0  
; SEQ ID NO 1  
; LENGTH: 412  
; TYPE: PRT  
; ORGANISM: Pyrococcus furiosus  
US-09-445-472-1

Query Match 44.1%; Score 37.5; DB 2; Length 412;  
Best Local Similarity 72.7%; Pred. No. 3e+02;  
Matches 8; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 3 NGESEKPYDDH 13  
Db 56 NGRS-YPYDDH 65

```
RESULT 49
US-10-090-624-1
; Sequence 1, Application US/10090624
; Patent No. 6783970
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-10-090-624-1

Query Match      44.1%; Score 37.5; DB 2; Length 412;
Best Local Similarity 72.7%; Pred. No. 3e+02;
Matches      8; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY      3 NGESEKPYDDH 13
Db      56 NGRS-YPYDDH 65

RESULT 50
US-08-894-818B-3
; Sequence 3, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; OTHER INFORMATION: /note= Xaa at position 428 is Gly or Val.
US-08-894-818B-3
```

```
Query Match      44.1%; Score 37.5; DB 2; Length 522;
Best Local Similarity 72.7%; Pred. No. 4e+02;
Matches      8; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
```

```
QY      3 NGESEKPYDDH 13
Db      56 NGRS-YPYDDH 65
```

```
Search completed: February 10, 2006, 23:36:19
Job time : 68 secs
```

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:47:05 ; Search time 180 Seconds  
(without alignments)  
32.498 Million cell updates/sec

Title: US-10-797-821-28

Perfect score: 85

Sequence: 1 QWNGSEKPYDDHL 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database :

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*

2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*

3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*

4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*

5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*

6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85	100.0	14	4	US-10-383-930-28
2	85	100.0	14	5	US-10-797-821-28
3	85	100.0	15	5	US-10-797-302-4
4	85	100.0	15	5	US-10-797-302-8
5	85	100.0	15	5	US-10-797-302-9
6	85	100.0	1590	4	US-10-383-930-37
7	85	100.0	1590	5	US-10-797-821-37
8	67	78.8	15	5	US-10-797-302-6
9	67	78.8	1375	3	US-09-740-274-4
10	67	78.8	1375	4	US-10-383-930-35
11	67	78.8	1375	5	US-10-797-821-35
12	67	78.8	1475	3	US-09-740-274-2
13	67	78.8	1475	4	US-10-383-930-34
14	67	78.8	1475	5	US-10-797-821-34
15	64	75.3	15	5	US-10-797-302-5
16	48	56.5	105	4	US-10-282-122A-78042
17	47	55.3	105	4	US-10-282-122A-68515
18	47	55.3	1006	5	US-10-484-218-22
19	46	54.1	504	4	US-10-724-972A-6956
20	46	54.1	864	4	US-10-437-963-15309
21	45	52.9	59	4	US-10-424-599-151656
22	45	52.9	105	4	US-10-282-122A-77550
23	45	52.9	178	4	US-10-424-599-262066
24	44	51.8	124	4	US-10-425-115-345898
25	44	51.8	175	4	US-10-156-761-9453
26	44	51.8	425	4	US-10-425-115-210862
27	44	51.8	480	4	US-10-425-114-65670

28	44	51.8	2447	4	US-10-130-115-28	Sequence 28, Appl
29	44	51.8	2447	4	US-10-369-072-28	Sequence 28, Appl
30	43	50.6	114	4	US-10-437-963-127860	Sequence 127860,
31	43	50.6	142	4	US-10-437-963-127805	Sequence 127805,
32	43	50.6	422	4	US-10-437-963-195318	Sequence 195318,
33	43	50.6	545	4	US-10-437-963-170027	Sequence 170027,
34	43	50.6	700	4	US-10-437-963-127893	Sequence 127893,
35	42.5	50.0	268	6	US-11-097-143-29574	Sequence 29574, A
36	42	49.4	105	4	US-10-282-122A-43094	Sequence 43094, A
37	42	49.4	105	4	US-10-282-122A-56312	Sequence 56312, A
38	42	49.4	105	4	US-10-282-122A-59859	Sequence 59859, A
39	42	49.4	105	4	US-10-282-122A-73308	Sequence 73308, A
40	42	49.4	105	4	US-10-282-122A-76003	Sequence 76003, A
41	42	49.4	188	5	US-10-450-763-34912	Sequence 34912, A
42	42	49.4	299	5	US-10-450-763-60378	Sequence 60378, A
43	42	49.4	424	3	US-09-815-242-13466	Sequence 13466, A
44	42	49.4	424	4	US-10-282-122A-74130	Sequence 74130, A
45	42	49.4	424	4	US-10-041-018-171	Sequence 171, App
46	42	49.4	424	4	US-10-041-018-270	Sequence 270, App
47	42	49.4	424	5	US-10-472-928-3562	Sequence 3562, Ap
48	42	49.4	426	5	US-10-617-320-3690	Sequence 3690, Ap
49	42	49.4	492	4	US-10-425-115-240837	Sequence 240837,
50	42	49.4	785	4	US-10-437-963-201769	Sequence 201769,
51	42	49.4	1439	4	US-10-369-493-5819	Sequence 5819, Ap
52	41	48.2	225	4	US-10-156-761-10051	Sequence 10051, A
53	41	48.2	304	4	US-10-112-944-806	Sequence 806, App
54	41	48.2	471	4	US-10-282-122A-68665	Sequence 68665, A
55	41	48.2	488	4	US-10-369-493-3680	Sequence 3680, Ap
56	41	48.2	546	4	US-10-156-761-10831	Sequence 10831, A
57	41	48.2	1202	4	US-10-112-944-352	Sequence 352, App
58	40.5	47.6	1601	4	US-10-437-963-195646	Sequence 195646,
59	40	47.1	40	4	US-10-424-599-179567	Sequence 179567,
60	40	47.1	199	4	US-10-437-963-160086	Sequence 160086,
61	40	47.1	229	4	US-10-424-599-273966	Sequence 273966,
62	40	47.1	238	4	US-10-425-114-68447	Sequence 68447, A
63	40	47.1	242	5	US-10-450-763-36299	Sequence 36299, A
64	40	47.1	300	4	US-10-424-599-194741	Sequence 194741,
65	40	47.1	304	4	US-10-425-114-36854	Sequence 36854, A
66	40	47.1	333	4	US-10-424-599-194752	Sequence 194752,
67	40	47.1	416	4	US-10-425-114-52792	Sequence 52792, A
68	40	47.1	423	4	US-10-425-114-59058	Sequence 59058, A
69	40	47.1	423	4	US-10-425-114-69404	Sequence 69404, A
70	40	47.1	436	4	US-10-282-122A-70602	Sequence 70602, A
71	40	47.1	438	4	US-10-767-701-44829	Sequence 44829, A
72	40	47.1	439	4	US-10-437-963-184446	Sequence 184446,
73	40	47.1	440	4	US-10-425-115-261600	Sequence 261600,
74	40	47.1	440	4	US-10-425-115-261602	Sequence 261602,
75	40	47.1	443	4	US-10-724-972A-4158	Sequence 4158, Ap
76	40	47.1	466	5	US-10-739-930-10930	Sequence 10930, A
77	40	47.1	474	4	US-10-425-114-65629	Sequence 65629, A
78	40	47.1	513	6	US-11-097-143-38688	Sequence 38688, A
79	40	47.1	532	3	US-09-801-368-360	Sequence 360, App
80	40	47.1	554	4	US-10-282-122A-68068	Sequence 68068, A
81	40	47.1	554	4	US-10-282-122A-68085	Sequence 68085, A
82	40	47.1	554	4	US-10-424-599-245422	Sequence 245422,
83	40	47.1	604	4	US-10-424-599-142871	Sequence 142871,
84	40	47.1	606	4	US-10-425-114-37181	Sequence 37181, A
85	40	47.1	1001	4	US-10-369-493-22344	Sequence 22344, A
86	40	47.1	1006	4	US-10-369-493-4198	Sequence 4198, Ap
87	40	47.1	1132	4	US-10-437-963-181220	Sequence 181220, A
88	40	47.1	2257	6	US-11-097-143-12138	Sequence 12138, A
89	39.5	46.5	58	3	US-09-864-761-44097	Sequence 44097, A
90	39.5	46.5	194	4	US-10-425-114-41983	Sequence 41983, A
91	39.5	46.5	414	5	US-10-916-064-4	Sequence 4, Appli
92	39.5	46.5	465	5	US-10-450-763-50549	Sequence 50549, A
93	39.5	46.5	1693	6	US-11-097-143-3573	Sequence 3573, Ap
94	39	45.9	91	4	US-10-424-599-202274	Sequence 202274,
95	39	45.9	101	4	US-10-425-115-305872	Sequence 305872,
96	39	45.9	103	4	US-10-767-701-62079	Sequence 62079, A
97	39	45.9	145	4	US-10-425-115-223874	Sequence 223874, A
98	39	45.9	157	4	US-10-282-122A-48297	Sequence 48297, A
99	39	45.9	177	4	US-10-424-599-276946	Sequence 276946,
100	39	45.9	199	4	US-10-425-115-223863	Sequence 223863,

101	39	45.9	203	4	US-10-492-100-10	Sequence 10, Appl	174	38	44.7	444	4	US-10-437-963-164896	Sequence 164896,
102	39	45.9	237	4	US-10-492-100-6	Sequence 6, Appl	175	38	44.7	445	4	US-10-097-111-289	Sequence 289, App
103	39	45.9	294	4	US-10-437-963-123992	Sequence 123992,	176	38	44.7	446	5	US-10-732-923-10254	Sequence 10254, A
104	39	45.9	313	4	US-10-724-372A-4754	Sequence 4754, Ap	177	38	44.7	509	4	US-10-094-749-2176	Sequence 2176, Ap
105	39	45.9	351	4	US-10-156-761-12036	Sequence 12036, A	178	38	44.7	511	4	US-10-425-115-367402	Sequence 367402,
106	39	45.9	360	4	US-10-087-464-12	Sequence 12, Appl	179	38	44.7	532	4	US-10-377-079-80	Sequence 80, Appl
107	39	45.9	366	4	US-10-323-069A-121	Sequence 121, App	180	38	44.7	599	4	US-10-287-218-11	Sequence 11, Appl
108	39	45.9	366	5	US-10-851-965-121	Sequence 121, App	181	38	44.7	599	4	US-10-108-260A-4013	Sequence 4013, Ap
109	39	45.9	427	4	US-10-437-963-200065	Sequence 20065, A	182	38	44.7	599	4	US-10-474-291-11	Sequence 11, Appl
110	39	45.9	432	5	US-10-437-963-38354	Sequence 38354, A	183	38	44.7	640	4	US-10-169-493-5226	Sequence 5226, Ap
111	39	45.9	548	5	US-10-491-472-13	Sequence 13, Appl	184	38	44.7	669	5	US-10-450-763-56601	Sequence 56601, A
112	39	45.9	561	4	US-10-369-493-12412	Sequence 12412, A	185	38	44.7	1010	4	US-10-369-493-3068	Sequence 3068, Ap
113	39	45.9	566	4	US-10-369-493-12412	Sequence 12412, A	186	38	44.7	1018	4	US-10-282-122A-63308	Sequence 63308, A
114	39	45.9	622	4	US-10-424-599-223097	Sequence 223097,	187	38	44.7	1021	4	US-10-437-963-121263	Sequence 121263,
115	39	45.9	622	4	US-10-425-114-51780	Sequence 51780, A	188	38	44.7	1033	4	US-10-437-963-121733	Sequence 121733,
116	39	45.9	648	4	US-10-369-493-11453	Sequence 11453, A	189	38	44.7	1095	4	US-10-262-794A-34	Sequence 34, Appl
117	39	45.9	671	4	US-10-369-493-8399	Sequence 8399, Ap	190	38	44.7	1172	4	US-10-437-963-121659	Sequence 121659,
118	39	45.9	1145	5	US-10-415-656-15	Sequence 15, Appl	191	38	44.7	1312	6	US-11-097-143-9057	Sequence 9057, Ap
119	39	45.9	1396	4	US-10-437-963-182927	Sequence 182927,	192	38	44.7	1934	4	US-10-437-963-204616	Sequence 204616,
120	39	45.9	1602	5	US-10-954-924-7	Sequence 7, Appl	193	38	44.7	2399	5	US-10-732-923-15036	Sequence 15036, A
121	39	45.9	1621	5	US-10-954-924-5	Sequence 5, Appl	194	37.5	44.1	79	4	US-10-424-599-200768	Sequence 200768,
122	39	45.9	1639	5	US-10-087-464-10	Sequence 10, Appl	195	37.5	44.1	223	4	US-10-425-115-237323	Sequence 237323,
123	39	45.9	1639	5	US-10-954-924-3	Sequence 3, Appl	196	37.5	44.1	229	4	US-10-424-599-195718	Sequence 195718,
124	39	45.9	2327	4	US-10-016-248-55	Sequence 55, Appl	197	37.5	44.1	233	4	US-10-424-599-195696	Sequence 195696,
125	38.5	45.3	452	6	US-11-046-124-8	Sequence 8, Appl	198	37.5	44.1	238	4	US-10-425-114-68829	Sequence 68829, A
126	38	44.7	52	4	US-10-425-115-326298	Sequence 326298,	199	37.5	44.1	321	4	US-10-424-599-145544	Sequence 145544,
127	38	44.7	62	4	US-10-424-599-211866	Sequence 211866,	200	37.5	44.1	332	4	US-10-166-225A-53	Sequence 53, Appl
128	38	44.7	67	4	US-10-425-115-185395	Sequence 185395,	201	37.5	44.1	366	4	US-10-424-599-195707	Sequence 195707,
129	38	44.7	69	4	US-10-029-386-28272	Sequence 28272, A	202	37.5	44.1	412	4	US-10-090-624-1	Sequence 1, Appl
130	38	44.7	78	4	US-10-425-115-190711	Sequence 190711,	203	37.5	44.1	412	5	US-10-888-588-1	Sequence 1, Appl
131	38	44.7	104	4	US-10-425-115-251568	Sequence 251568,	204	37.5	44.1	432	3	US-10-437-963-110279	Sequence 110279,
132	38	44.7	104	4	US-10-425-115-274178	Sequence 274178,	205	37.5	44.1	522	3	US-09-841-553-3	Sequence 3, Appl
133	38	44.7	105	4	US-10-424-599-232000	Sequence 232000,	206	37.5	44.1	522	4	US-10-090-624-4	Sequence 4, Appl
134	38	44.7	108	5	US-10-439-353A-486	Sequence 486, App	207	37.5	44.1	522	5	US-10-800-684-3	Sequence 3, Appl
135	38	44.7	108	5	US-10-499-353A-487	Sequence 487, App	208	37.5	44.1	522	5	US-10-888-588-4	Sequence 4, Appl
136	38	44.7	112	4	US-10-424-599-180410	Sequence 180410,	209	37.5	44.1	654	3	US-09-841-553-35	Sequence 35, Appl
137	38	44.7	118	4	US-10-425-115-228403	Sequence 228403,	210	37.5	44.1	654	4	US-10-090-624-16	Sequence 16, Appl
138	38	44.7	124	4	US-10-767-701-43550	Sequence 43550, A	211	37.5	44.1	654	5	US-10-800-684-35	Sequence 35, Appl
139	38	44.7	125	4	US-10-424-599-209379	Sequence 209379,	212	37.5	44.1	654	5	US-10-888-588-16	Sequence 16, Appl
140	38	44.7	137	4	US-10-424-599-250482	Sequence 250482,	213	37.5	44.1	659	2	US-09-841-553-5	Sequence 5, Appl
141	38	44.7	156	5	US-10-732-923-21528	Sequence 21528, A	214	37.5	44.1	659	5	US-10-800-684-5	Sequence 5, Appl
142	38	44.7	156	5	US-10-732-923-21529	Sequence 21529, A	215	37.5	44.1	809	4	US-11-097-143-2370	Sequence 311, App
143	38	44.7	162	4	US-10-425-115-292979	Sequence 292979,	216	37.5	44.1	882	6	US-10-108-605-311	Sequence 50935, A
144	38	44.7	170	4	US-10-425-114-43920	Sequence 43920, A	217	37.5	44.1	936	5	US-10-450-763-50935	Sequence 20747,
145	38	44.7	175	4	US-10-424-599-219855	Sequence 219855,	218	37	43.5	36	4	US-10-424-599-220747	Sequence 220747,
146	38	44.7	186	4	US-10-174-695-4	Sequence 4, Appl	219	37	43.5	59	4	US-10-425-115-336681	Sequence 336681,
147	38	44.7	198	4	US-10-282-122A-51498	Sequence 51498, A	220	37	43.5	61	4	US-10-424-599-14760	Sequence 14760,
148	38	44.7	207	4	US-10-424-599-232005	Sequence 232005,	221	37	43.5	69	4	US-10-425-115-211361	Sequence 211361,
149	38	44.7	207	4	US-10-425-114-45766	Sequence 45766, A	222	37	43.5	70	3	US-09-864-761-41852	Sequence 41852, A
150	38	44.7	222	4	US-10-369-493-13213	Sequence 13213, A	223	37	43.5	74	4	US-10-437-963-108821	Sequence 108821,
151	38	44.7	238	4	US-10-424-599-244722	Sequence 244722,	224	37	43.5	88	4	US-10-425-115-325442	Sequence 325442,
152	38	44.7	254	4	US-10-424-599-158554	Sequence 158554,	225	37	43.5	101	4	US-10-029-386-268813	Sequence 268813, A
153	38	44.7	258	4	US-10-425-114-63671	Sequence 63671, A	226	37	43.5	101	4	US-10-437-963-168485	Sequence 168485,
154	38	44.7	258	4	US-10-425-115-287733	Sequence 287733,	227	37	43.5	104	5	US-10-450-763-57208	Sequence 57208, A
155	38	44.7	274	4	US-10-424-599-158555	Sequence 158555,	228	37	43.5	111	4	US-10-424-599-196899	Sequence 196899,
156	38	44.7	274	5	US-10-739-930-9431	Sequence 9431, Ap	229	37	43.5	118	4	US-10-425-115-222601	Sequence 222601,
157	38	44.7	282	4	US-10-437-963-122500	Sequence 122500,	230	37	43.5	119	4	US-10-424-599-165566	Sequence 165566,
158	38	44.7	296	4	US-10-424-599-232010	Sequence 232010,	231	37	43.5	121	4	US-10-437-963-149812	Sequence 149812,
159	38	44.7	316	4	US-10-425-115-287737	Sequence 287737,	232	37	43.5	123	4	US-10-767-701-52675	Sequence 52675, A
160	38	44.7	335	4	US-10-425-115-347367	Sequence 347367,	233	37	43.5	126	3	US-09-864-761-36479	Sequence 36479, A
161	38	44.7	380	3	US-09-998-279-26	Sequence 26, Appl	234	37	43.5	135	4	US-10-425-115-328828	Sequence 328828,
162	38	44.7	380	4	US-10-283-024-4	Sequence 4, Appl	235	37	43.5	138	4	US-10-424-599-219857	Sequence 219857,
163	38	44.7	385	3	US-09-998-279-24	Sequence 24, Appl	236	37	43.5	147	5	US-10-450-763-31486	Sequence 31486, A
164	38	44.7	391	4	US-10-424-599-232009	Sequence 232009,	237	37	43.5	151	4	US-10-767-701-59911	Sequence 59911, A
165	38	44.7	393	4	US-10-437-963-132805	Sequence 132805,	238	37	43.5	161	4	US-10-437-963-31822	Sequence 31822, A
166	38	44.7	403	4	US-10-425-114-59200	Sequence 59200, A	239	37	43.5	170	4	US-10-424-599-221347	Sequence 221347,
167	38	44.7	408	4	US-10-425-114-55605	Sequence 55605, A	240	37	43.5	172	4	US-10-282-122A-72439	Sequence 72439, A
168	38	44.7	408	4	US-10-425-114-56199	Sequence 56199, A	241	37	43.5	184	4	US-10-425-115-243718	Sequence 243718,
169	38	44.7	412	4	US-10-437-963-151267	Sequence 151267,	242	37	43.5	215	4	US-10-425-115-348837	Sequence 348837,
170	38	44.7	413	4	US-10-032-585-7066	Sequence 7066, Ap	243	37	43.5	228	4	US-10-437-963-160205	Sequence 160205,
171	38	44.7	418	4	US-10-424-599-229160	Sequence 229160,	244	37	43.5	229	4	US-10-425-114-69102	Sequence 69102, A
172	38	44.7	423	4	US-10-425-115-287739	Sequence 287739,	245	37	43.5	233	4	US-10-425-114-50012	Sequence 50012, A
173	38	44.7	441	4	US-10-424-599-232008	Sequence 232008,	246	37	43.5	251	4	US-10-425-114-50012	Sequence 50012, A





393	37	43.5	1889	4	US-10-437-963-149886	Sequence 149886,	466	37	43.5	2055	4	US-10-437-963-203325	Sequence 203325,
394	37	43.5	1892	4	US-10-437-963-166972	Sequence 166972,	467	37	43.5	2131	4	US-10-437-963-181289	Sequence 181289,
395	37	43.5	1894	4	US-10-437-963-149938	Sequence 149938,	468	37	43.5	2144	5	US-10-723-860-2303	Sequence 2303, Ap
396	37	43.5	1899	4	US-10-437-963-203285	Sequence 203285,	469	37	43.5	2170	4	US-10-437-963-149857	Sequence 149857,
397	37	43.5	1902	4	US-10-437-963-181407	Sequence 181407,	470	37	43.5	2289	4	US-10-437-963-149715	Sequence 149715,
398	37	43.5	1902	4	US-10-437-963-203253	Sequence 203253,	471	37	43.5	2251	4	US-10-144-194A-80	Sequence 80, Appl
399	37	43.5	1911	4	US-10-437-963-149855	Sequence 149855,	472	37	43.5	2251	5	US-10-491-566-80	Sequence 82, Appl
400	37	43.5	1913	4	US-10-437-963-184866	Sequence 184866,	473	37	43.5	2633	4	US-10-144-194A-82	Sequence 82, Appl
401	37	43.5	1915	4	US-10-408-765A-2936	Sequence 2936, Ap	474	37	43.5	2633	5	US-10-491-566-82	Sequence 82, Appl
402	37	43.5	1928	4	US-10-437-963-149971	Sequence 149971,	475	37	43.5	2724	3	US-09-808-602-13	Sequence 13, Appl
403	37	43.5	1930	4	US-10-437-963-181294	Sequence 181294,	476	37	43.5	2724	3	US-09-800-198-13	Sequence 13, Appl
404	37	43.5	1937	4	US-10-437-963-149883	Sequence 149883,	477	37	43.5	2733	3	US-09-808-602-8	Sequence 8, Appl
405	37	43.5	1942	4	US-10-437-963-149709	Sequence 149709,	478	37	43.5	2733	3	US-09-800-198-8	Sequence 8, Appl
406	37	43.5	1945	4	US-10-437-963-149932	Sequence 149932,	479	37	43.5	2759	4	US-10-072-012-144	Sequence 144, Ap
407	37	43.5	1951	4	US-10-424-599-230751	Sequence 230751,	480	37	43.5	2759	4	US-09-808-602-80	Sequence 80, Appl
408	37	43.5	1955	4	US-10-437-963-149819	Sequence 149819,	481	37	43.5	2764	3	US-09-800-198-68	Sequence 68, Appl
409	37	43.5	1956	4	US-10-437-963-149712	Sequence 149712,	482	37	43.5	2764	4	US-10-072-012-487	Sequence 487, App
410	37	43.5	1963	4	US-10-437-963-149901	Sequence 149901,	483	37	43.5	2764	5	US-10-631-467-1514	Sequence 1514, App
411	37	43.5	1964	4	US-10-437-963-149709	Sequence 149709,	484	37	43.5	2765	3	US-09-808-602-84	Sequence 84, Appl
412	37	43.5	1964	4	US-10-437-963-149929	Sequence 149929,	485	37	43.5	2765	3	US-09-800-198-72	Sequence 72, Appl
413	37	43.5	1964	4	US-10-437-963-149931	Sequence 149931,	486	37	43.5	2765	4	US-10-072-012-488	Sequence 488, App
414	37	43.5	1964	4	US-10-437-963-203321	Sequence 203321,	487	37	43.5	2802	3	US-09-808-602-81	Sequence 81, Appl
415	37	43.5	1966	4	US-10-437-963-203209	Sequence 203209,	488	37	43.5	2802	3	US-09-800-198-69	Sequence 69, Appl
416	37	43.5	1967	4	US-10-219-834-85	Sequence 85, Appl	489	37	43.5	2802	4	US-10-072-012-489	Sequence 489, App
417	37	43.5	1967	4	US-10-325-567A-575	Sequence 575, App	490	37	43.5	2802	4	US-10-437-963-149734	Sequence 149734,
418	37	43.5	1968	4	US-10-437-963-149862	Sequence 149862,	491	37	43.5	3530	4	US-10-296-115-1101	Sequence 1101, Ap
419	37	43.5	1968	4	US-10-437-963-182423	Sequence 182423,	492	37	43.5	4131	4	US-10-369-493-5136	Sequence 5136, Ap
420	37	43.5	1969	4	US-10-437-963-149823	Sequence 149823,	493	37	43.5	6304	4	US-10-147-026-16	Sequence 16, Appl
421	37	43.5	1971	4	US-10-437-963-149981	Sequence 149981,	494	36.5	42.9	54	4	US-10-424-599-156436	Sequence 156436,
422	37	43.5	1972	4	US-10-437-963-185612	Sequence 185612,	495	36.5	42.9	75	4	US-10-425-115-260466	Sequence 260466,
423	37	43.5	1974	4	US-10-437-963-203173	Sequence 203173,	496	36.5	42.9	97	4	US-10-424-599-237861	Sequence 237861,
424	37	43.5	1975	4	US-10-437-963-149771	Sequence 149771,	497	36.5	42.9	150	5	US-10-501-282-3928	Sequence 3928, Ap
425	37	43.5	1975	4	US-10-437-963-149782	Sequence 149782,	498	36.5	42.9	179	3	US-09-872-523-59	Sequence 59, Appl
426	37	43.5	1975	4	US-10-437-963-203183	Sequence 203183,	499	36.5	42.9	179	5	US-10-839-896-59	Sequence 59, Appl
427	37	43.5	1979	4	US-10-437-963-149933	Sequence 149933,	500	36.5	42.9	187	3	US-09-872-523-61	Sequence 61, Appl
428	37	43.5	1982	4	US-10-437-963-149741	Sequence 149741,	501	36.5	42.9	187	5	US-10-839-896-61	Sequence 61, Appl
429	37	43.5	1984	4	US-10-437-963-203182	Sequence 203182,	502	36.5	42.9	207	5	US-10-501-282-3930	Sequence 3930, Ap
430	37	43.5	1986	4	US-10-437-963-149708	Sequence 149708,	503	36.5	42.9	228	4	US-10-425-114-55459	Sequence 55459, A
431	37	43.5	1986	4	US-10-437-963-149780	Sequence 149780,	504	36.5	42.9	228	4	US-10-971-144-44	Sequence 44, Appl
432	37	43.5	1986	4	US-10-437-963-203212	Sequence 203212,	505	36.5	42.9	276	5	US-10-971-144-34	Sequence 34, Appl
433	37	43.5	1986	4	US-10-437-963-203323	Sequence 203323,	506	36.5	42.9	293	5	US-10-971-144-30	Sequence 30, Appl
434	37	43.5	1988	4	US-10-437-963-149976	Sequence 149976,	507	36.5	42.9	296	4	US-10-768-976-48	Sequence 48, Appl
435	37	43.5	1989	4	US-10-437-963-203168	Sequence 203168,	508	36.5	42.9	352	5	US-10-450-763-54735	Sequence 54735, A
436	37	43.5	1991	4	US-10-437-963-203207	Sequence 203207,	509	36.5	42.9	446	5	US-10-971-144-32	Sequence 32, Appl
437	37	43.5	1991	4	US-10-437-963-203284	Sequence 203284,	510	36.5	42.9	672	4	US-10-437-963-181175	Sequence 181175,
438	37	43.5	1991	4	US-10-437-963-203286	Sequence 203286,	511	36.5	42.9	851	4	US-10-437-963-195578	Sequence 195578,
439	37	43.5	1992	4	US-10-437-963-149983	Sequence 149983,	512	36.5	42.9	1032	4	US-10-282-122A-42928	Sequence 42928, A
440	37	43.5	1995	4	US-10-437-963-149740	Sequence 149740,	513	36.5	42.9	1108	4	US-10-437-963-118940	Sequence 118940,
441	37	43.5	1995	4	US-10-437-963-149753	Sequence 149753,	514	36.5	42.9	2072	4	US-10-437-963-195535	Sequence 195535,
442	37	43.5	1995	4	US-10-437-963-203244	Sequence 203244,	515	36	42.4	46	4	US-10-424-599-243577	Sequence 243577,
443	37	43.5	2001	4	US-10-437-963-149972	Sequence 149972,	516	36	42.4	48	4	US-10-425-115-250069	Sequence 250069,
444	37	43.5	2001	4	US-10-437-963-182972	Sequence 182972,	517	36	42.4	55	4	US-10-767-701-50276	Sequence 50276, A
445	37	43.5	2002	4	US-10-437-963-149896	Sequence 149896,	518	36	42.4	57	3	US-09-864-763-38577	Sequence 38577, A
446	37	43.5	2004	4	US-10-437-963-203293	Sequence 203293,	519	36	42.4	58	4	US-10-029-386-31700	Sequence 31700, A
447	37	43.5	2005	4	US-10-437-963-149706	Sequence 149706,	520	36	42.4	70	4	US-10-767-701-47314	Sequence 47314, A
448	37	43.5	2005	4	US-10-437-963-149706	Sequence 203219	521	36	42.4	73	3	US-09-738-626-6469	Sequence 6469, Ap
449	37	43.5	2008	4	US-10-437-963-149851	Sequence 149851,	522	36	42.4	73	4	US-10-425-115-235505	Sequence 235505,
450	37	43.5	2008	4	US-10-437-963-182237	Sequence 182237,	523	36	42.4	73	5	US-10-494-672-418	Sequence 418, App
451	37	43.5	2008	4	US-10-437-963-184738	Sequence 184738,	524	36	42.4	86	4	US-10-424-599-224162	Sequence 224162,
452	37	43.5	2010	4	US-10-437-963-149775	Sequence 149775,	525	36	42.4	87	4	US-10-106-698-6122	Sequence 6122, Ap
453	37	43.5	2013	4	US-10-437-963-149974	Sequence 149974,	526	36	42.4	96	4	US-10-425-115-293371	Sequence 293371,
454	37	43.5	2013	4	US-10-437-963-203324	Sequence 203324,	527	36	42.4	99	4	US-10-767-701-37892	Sequence 37892, A
455	37	43.5	2014	4	US-10-437-963-182499	Sequence 182499,	528	36	42.4	105	4	US-10-282-122A-59128	Sequence 59128, A
456	37	43.5	2016	4	US-10-437-963-149867	Sequence 149867,	529	36	42.4	105	4	US-10-282-122A-67243	Sequence 67243, A
457	37	43.5	2016	4	US-10-437-963-184864	Sequence 184864,	530	36	42.4	109	4	US-10-078-929-208	Sequence 208, App
458	37	43.5	2018	4	US-10-437-963-181416	Sequence 181416,	531	36	42.4	109	4	US-10-408-765A-1912	Sequence 1912, Ap
459	37	43.5	2018	4	US-10-437-963-184861	Sequence 184861,	532	36	42.4	111	4	US-10-437-963-128992	Sequence 128992,
460	37	43.5	2020	4	US-10-437-963-183621	Sequence 183621,	533	36	42.4	120	4	US-10-425-114-70374	Sequence 70374, A
461	37	43.5	2021	4	US-10-437-963-167916	Sequence 167916,	534	36	42.4	120	4	US-10-425-115-209687	Sequence 209687,
462	37	43.5	2021	4	US-10-437-963-181530	Sequence 181530,	535	36	42.4	124	3	US-09-997-807-10	Sequence 10, Appl
463	37	43.5	2021	4	US-10-437-963-182553	Sequence 182553,	536	36	42.4	124	4	US-10-370-370-10	Sequence 10, Appl
464	37	43.5	2021	4	US-10-437-963-203243	Sequence 203243,	537	36	42.4	124	6	US-11-046-203-10	Sequence 10, Appl
465	37	43.5	2048	4	US-10-437-963-183905	Sequence 183905,	538	36	42.4	130	4	US-10-767-701-55428	Sequence 55428, A

539	36	42.4	134	4	US-10-437-963-193678	Sequence 193678,	612	36	42.4	357	4	US-10-425-115-243466	Sequence 243466,
540	36	42.4	138	4	US-10-424-599-189170	Sequence 189170,	613	36	42.4	375	4	US-10-425-115-243466	Sequence 123, App
541	36	42.4	138	4	US-10-424-599-253347	Sequence 253347,	613	36	42.4	375	5	US-10-851-965-123	Sequence 123, App
542	36	42.4	139	4	US-10-078-929-94	Sequence 94, Appl	615	36	42.4	380	4	US-10-437-963-168875	Sequence 168875,
543	36	42.4	140	4	US-10-437-963-202506	Sequence 202506,	616	36	42.4	396	4	US-10-437-963-168875	Sequence 43325, A
544	36	42.4	156	4	US-10-424-599-245469	Sequence 245469,	617	36	42.4	404	4	US-10-437-963-169199	Sequence 169199,
545	36	42.4	161	4	US-10-078-929-96	Sequence 96, Appl	618	36	42.4	410	4	US-10-369-493-18556	Sequence 18556, A
546	36	42.4	161	4	US-10-424-599-145839	Sequence 145839,	619	36	42.4	415	4	US-10-389-566-2345	Sequence 2345, Ap
547	36	42.4	176	4	US-10-424-599-156922	Sequence 156922,	620	36	42.4	415	5	US-10-732-923-17924	Sequence 17924, A
548	36	42.4	177	4	US-10-425-115-210981	Sequence 210981,	621	36	42.4	425	4	US-10-369-493-9687	Sequence 9687, Ap
549	36	42.4	178	3	US-09-997-807-6	Sequence 6, Appli	622	36	42.4	428	3	US-09-792-420-8	Sequence 8, Appli
550	36	42.4	178	4	US-10-370-370-6	Sequence 6, Appli	623	36	42.4	434	4	US-10-282-122A-69137	Sequence 69137, A
551	36	42.4	178	6	US-11-046-203-6	Sequence 6, Appli	624	36	42.4	436	5	US-10-739-930-8146	Sequence 8146, Ap
552	36	42.4	181	4	US-10-284-049-4022	Sequence 4022, Ap	625	36	42.4	447	5	US-10-875-127-53	Sequence 53, Appl
553	36	42.4	181	5	US-10-450-763-48890	Sequence 48890, Ap	626	36	42.4	448	5	US-10-501-282-2386	Sequence 2386, Ap
554	36	42.4	181	4	US-10-369-493-21981	Sequence 21981,	627	36	42.4	453	3	US-09-764-875-752	Sequence 752, App
555	36	42.4	183	5	US-10-450-763-37035	Sequence 37035, A	628	36	42.4	462	4	US-10-437-963-204495	Sequence 204495,
556	36	42.4	189	4	US-10-425-115-210980	Sequence 210980,	629	36	42.4	463	4	US-10-425-115-317759	Sequence 317759,
557	36	42.4	192	3	US-09-796-692-1790	Sequence 1790, Ap	630	36	42.4	464	4	US-10-437-963-121243	Sequence 121243,
558	36	42.4	192	3	US-09-796-692-2054	Sequence 2054, Ap	631	36	42.4	464	4	US-10-437-963-162952	Sequence 162952,
559	36	42.4	192	4	US-10-040-862-1790	Sequence 1790, Ap	632	36	42.4	473	4	US-10-437-963-184973	Sequence 184973,
560	36	42.4	192	4	US-10-040-862-2054	Sequence 2054, Ap	633	36	42.4	474	4	US-10-425-114-37484	Sequence 37484, A
561	36	42.4	192	4	US-10-057-475B-1790	Sequence 1790, Ap	634	36	42.4	481	5	US-10-501-282-2388	Sequence 2388, Ap
562	36	42.4	192	4	US-10-057-475B-2054	Sequence 2054, Ap	635	36	42.4	482	5	US-10-450-763-59048	Sequence 59048, A
563	36	42.4	192	4	US-10-354-884B-1790	Sequence 1790, Ap	636	36	42.4	493	5	US-10-501-282-2390	Sequence 2390, Ap
564	36	42.4	192	4	US-10-354-884B-2054	Sequence 2054, Ap	637	36	42.4	504	4	US-10-479-435-21	Sequence 21, Appl
565	36	42.4	192	4	US-10-764-324-1790	Sequence 1790, Ap	638	36	42.4	512	5	US-10-501-282-2392	Sequence 2392, Ap
566	36	42.4	192	4	US-10-764-324-2054	Sequence 2054, Ap	639	36	42.4	516	4	US-10-369-493-13111	Sequence 13111, A
567	36	42.4	193	4	US-10-425-114-53103	Sequence 53103, A	640	36	42.4	520	4	US-10-282-122A-62068	Sequence 62068, A
568	36	42.4	202	4	US-10-029-386-33717	Sequence 33717, A	641	36	42.4	520	6	US-11-097-143-29943	Sequence 29943, A
569	36	42.4	221	5	US-10-484-703-37	Sequence 37, Appl	642	36	42.4	525	6	US-11-097-143-12015	Sequence 12015, A
570	36	42.4	223	4	US-10-424-599-237042	Sequence 237042,	643	36	42.4	527	4	US-10-424-599-162222	Sequence 162222,
571	36	42.4	225	4	US-10-424-599-156920	Sequence 156920,	644	36	42.4	532	5	US-10-950-095-28	Sequence 28, Appl
572	36	42.4	225	4	US-10-424-599-200284	Sequence 200284,	645	36	42.4	545	4	US-10-437-963-148730	Sequence 148730,
573	36	42.4	230	4	US-10-424-599-270398	Sequence 270398,	646	36	42.4	545	5	US-10-450-763-47076	Sequence 47076, A
574	36	42.4	236	4	US-10-424-599-156921	Sequence 156921,	647	36	42.4	550	5	US-10-053-248-10	Sequence 10, Appl
575	36	42.4	240	5	US-10-732-923-7384	Sequence 7384, Ap	648	36	42.4	558	4	US-10-364-748-2	Sequence 2, Appli
576	36	42.4	243	4	US-10-437-963-134835	Sequence 134835,	649	36	42.4	558	4	US-10-345-837-10	Sequence 10, Appl
577	36	42.4	256	4	US-10-424-599-199754	Sequence 199754,	650	36	42.4	579	6	US-11-097-143-16530	Sequence 16530, A
578	36	42.4	264	4	US-10-425-115-332825	Sequence 332825,	651	36	42.4	592	4	US-10-437-963-148732	Sequence 148732,
579	36	42.4	269	3	US-09-809-920-11	Sequence 11, Appl	652	36	42.4	594	5	US-10-450-763-49527	Sequence 49527, A
580	36	42.4	275	4	US-10-425-114-44781	Sequence 44781, A	653	36	42.4	607	4	US-10-282-122A-62418	Sequence 62418, A
581	36	42.4	283	3	US-09-952-689-5	Sequence 5, Appl1	654	36	42.4	614	4	US-10-425-114-63833	Sequence 63833, A
582	36	42.4	283	4	US-10-437-963-125606	Sequence 125606,	655	36	42.4	616	4	US-10-291-172-218	Sequence 218, App
583	36	42.4	285	4	US-10-424-599-149423	Sequence 149423,	656	36	42.4	616	4	US-10-221-278-218	Sequence 218, App
584	36	42.4	285	4	US-10-451-467A-526	Sequence 526, App	657	36	42.4	624	4	US-10-282-122A-64877	Sequence 64877, A
585	36	42.4	294	4	US-10-425-115-210986	Sequence 210986,	658	36	42.4	626	4	US-10-029-386-32226	Sequence 32226, A
586	36	42.4	294	4	US-10-425-115-210987	Sequence 210987,	659	36	42.4	635	4	US-10-437-963-182698	Sequence 182698,
587	36	42.4	295	4	US-10-425-115-54820	Sequence 54820, A	660	36	42.4	644	4	US-10-325-298-654	Sequence 135994,
588	36	42.4	303	3	US-09-764-875-1055	Sequence 1055, Ap	661	36	42.4	663	3	US-10-102-806-654	Sequence 654, App
589	36	42.4	303	4	US-10-158-034-72	Sequence 72, Appl	662	36	42.4	663	4	US-10-408-765A-616	Sequence 616, App
590	36	42.4	307	4	US-09-864-761-3433	Sequence 3433, A	663	36	42.4	664	4	US-10-408-765A-617	Sequence 617, App
591	36	42.4	311	3	US-09-864-761-3433	Sequence 3433, A	664	36	42.4	664	4	US-10-408-765A-711	Sequence 711, App
592	36	42.4	317	3	US-09-815-242-13426	Sequence 13426, A	665	36	42.4	664	4	US-10-408-765A-2414	Sequence 2414, Ap
593	36	42.4	317	3	US-09-815-242-13669	Sequence 13669, A	666	36	42.4	673	4	US-10-425-114-69138	Sequence 69138, A
594	36	42.4	317	4	US-10-166-225A-85	Sequence 85, Appl	667	36	42.4	673	4	US-10-425-115-237171	Sequence 237171,
595	36	42.4	317	5	US-10-472-928-600	Sequence 600, App	668	36	42.4	678	4	US-10-425-115-288028	Sequence 288028,
596	36	42.4	319	6	US-11-097-143-17379	Sequence 17379, A	669	36	42.4	712	4	US-10-437-963-144606	Sequence 144606,
597	36	42.4	327	4	US-10-437-963-122985	Sequence 122985,	670	36	42.4	712	4	US-10-369-493-13792	Sequence 13792, A
598	36	42.4	336	5	US-10-739-930-10533	Sequence 10533, A	671	36	42.4	713	4	US-10-001-054-30	Sequence 30, Appl
599	36	42.4	340	4	US-09-864-761-35894	Sequence 35894, A	672	36	42.4	716	4	US-10-028-072-512	Sequence 512, App
600	36	42.4	340	4	US-10-767-701-38315	Sequence 38315, A	673	36	42.4	716	4	US-10-140-808-512	Sequence 512, App
601	36	42.4	344	4	US-10-282-122A-73789	Sequence 73789, A	674	36	42.4	716	4	US-10-121-049-512	Sequence 512, App
602	36	42.4	349	4	US-10-369-493-175	Sequence 175, App	675	36	42.4	716	4	US-10-123-904-512	Sequence 512, App
603	36	42.4	349	4	US-10-320-797-3125	Sequence 3125, Ap	676	36	42.4	716	4	US-10-140-470-512	Sequence 512, App
604	36	42.4	354	4	US-10-289-762-504	Sequence 504, App	677	36	42.4	716	4	US-10-175-746-512	Sequence 512, App
605	36	42.4	354	4	US-10-425-115-264894	Sequence 264894,	678	36	42.4	716	4	US-10-176-918-512	Sequence 512, App
606	36	42.4	358	4	US-10-389-566-2281	Sequence 2281, Ap	679	36	42.4	716	4	US-10-176-918-512	Sequence 512, App
607	36	42.4	358	5	US-10-732-923-17925	Sequence 17925, A	680	36	42.4	716	4	US-10-176-921-512	Sequence 512, App
608	36	42.4	359	5	US-10-617-320-2646	Sequence 2646, Ap	681	36	42.4	716	4	US-10-227-884-190	Sequence 190, App
609	36	42.4	360	4	US-10-437-963-201504	Sequence 201504,	682	36	42.4	716	4	US-10-137-865-512	Sequence 512, App
610	36	42.4	362	6	US-11-097-143-15438	Sequence 15438, A	683	36	42.4	716	4	US-10-140-474-512	Sequence 512, App
611	36	42.4	364	6	US-11-097-143-14841	Sequence 14841, A	684	36	42.4	716	4	US-10-142-431-512	Sequence 512, App

685	36	42.4	716	4	US-10-143-114-512	Sequence 512, App	758	36	42.4	716	4	US-10-127-843A-512	Sequence 512, App
686	36	42.4	716	4	US-10-230-163-190	Sequence 190, App	759	36	42.4	716	4	US-10-127-845A-512	Sequence 512, App
687	36	42.4	716	4	US-10-230-338-190	Sequence 190, App	760	36	42.4	716	4	US-10-127-846A-512	Sequence 512, App
688	36	42.4	716	4	US-10-142-419-512	Sequence 512, App	761	36	42.4	716	4	US-10-127-848A-512	Sequence 512, App
689	36	42.4	716	4	US-10-218-631-190	Sequence 190, App	762	36	42.4	716	4	US-10-127-849A-512	Sequence 512, App
690	36	42.4	716	4	US-10-123-362-512	Sequence 512, App	763	36	42.4	716	4	US-10-127-850A-512	Sequence 512, App
691	36	42.4	716	4	US-10-142-423-512	Sequence 512, App	764	36	42.4	716	4	US-10-127-851A-512	Sequence 512, App
692	36	42.4	716	4	US-10-230-414-190	Sequence 190, App	765	36	42.4	716	4	US-10-128-684A-512	Sequence 512, App
693	36	42.4	716	4	US-10-121-050-512	Sequence 512, App	766	36	42.4	716	4	US-10-128-686A-512	Sequence 512, App
694	36	42.4	716	4	US-10-141-755-512	Sequence 512, App	767	36	42.4	716	4	US-10-128-690A-512	Sequence 512, App
695	36	42.4	716	4	US-10-143-032-512	Sequence 512, App	768	36	42.4	716	4	US-10-128-691A-512	Sequence 512, App
696	36	42.4	716	4	US-10-232-224-190	Sequence 190, App	769	36	42.4	716	4	US-10-131-819A-512	Sequence 512, App
697	36	42.4	716	4	US-10-123-108-512	Sequence 512, App	770	36	42.4	716	4	US-10-131-829A-512	Sequence 512, App
698	36	42.4	716	4	US-10-123-236-512	Sequence 512, App	771	36	42.4	716	4	US-10-131-836A-512	Sequence 512, App
699	36	42.4	716	4	US-10-123-261-512	Sequence 512, App	772	36	42.4	716	4	US-10-146-729-512	Sequence 512, App
700	36	42.4	716	4	US-10-140-521-512	Sequence 512, App	773	36	42.4	716	4	US-10-146-791-512	Sequence 512, App
701	36	42.4	716	4	US-10-140-528-512	Sequence 512, App	774	36	42.4	716	4	US-10-147-484-512	Sequence 512, App
702	36	42.4	716	4	US-10-216-159A-190	Sequence 190, App	775	36	42.4	716	4	US-10-147-508-512	Sequence 512, App
703	36	42.4	716	4	US-10-121-045-512	Sequence 512, App	776	36	42.4	716	4	US-10-147-512-512	Sequence 512, App
704	36	42.4	716	4	US-10-123-292-512	Sequence 512, App	777	36	42.4	716	4	US-10-175-735-512	Sequence 512, App
705	36	42.4	716	4	US-10-123-303-512	Sequence 512, App	778	36	42.4	716	4	US-10-121-040-512	Sequence 512, App
706	36	42.4	716	4	US-10-124-819-512	Sequence 512, App	779	36	42.4	716	4	US-10-121-056-512	Sequence 512, App
707	36	42.4	716	4	US-10-124-822-512	Sequence 512, App	780	36	42.4	716	4	US-10-121-061-512	Sequence 512, App
708	36	42.4	716	4	US-10-140-925-512	Sequence 512, App	781	36	42.4	716	4	US-10-123-235-512	Sequence 512, App
709	36	42.4	716	4	US-10-160-498-512	Sequence 512, App	782	36	42.4	716	4	US-10-124-818-512	Sequence 512, App
710	36	42.4	716	4	US-10-218-849-190	Sequence 190, App	783	36	42.4	716	4	US-10-137-868-512	Sequence 512, App
711	36	42.4	716	4	US-10-227-873-190	Sequence 190, App	784	36	42.4	716	4	US-10-147-492-512	Sequence 512, App
712	36	42.4	716	4	US-10-227-883-190	Sequence 190, App	785	36	42.4	716	4	US-10-158-782-512	Sequence 512, App
713	36	42.4	716	4	US-10-124-824-512	Sequence 512, App	786	36	42.4	716	4	US-10-123-905-512	Sequence 512, App
714	36	42.4	716	4	US-10-127-825A-512	Sequence 512, App	787	36	42.4	716	4	US-10-123-907-512	Sequence 512, App
715	36	42.4	716	4	US-10-131-813A-512	Sequence 512, App	788	36	42.4	716	4	US-10-124-815-512	Sequence 512, App
716	36	42.4	716	4	US-10-127-829A-512	Sequence 512, App	789	36	42.4	716	4	US-10-125-921A-512	Sequence 512, App
717	36	42.4	716	4	US-10-127-835A-512	Sequence 512, App	790	36	42.4	716	4	US-10-125-928A-512	Sequence 512, App
718	36	42.4	716	4	US-10-127-839A-512	Sequence 512, App	791	36	42.4	716	4	US-10-127-821A-512	Sequence 512, App
719	36	42.4	716	4	US-10-127-801A-512	Sequence 512, App	792	36	42.4	716	4	US-10-127-822A-512	Sequence 512, App
720	36	42.4	716	4	US-10-128-693A-512	Sequence 512, App	793	36	42.4	716	4	US-10-127-824A-512	Sequence 512, App
721	36	42.4	716	4	US-10-131-813A-512	Sequence 512, App	794	36	42.4	716	4	US-10-127-826A-512	Sequence 512, App
722	36	42.4	716	4	US-10-131-818A-512	Sequence 512, App	795	36	42.4	716	4	US-10-127-827A-512	Sequence 512, App
723	36	42.4	716	4	US-10-131-823A-512	Sequence 512, App	796	36	42.4	716	4	US-10-127-828A-512	Sequence 512, App
724	36	42.4	716	4	US-10-131-824A-512	Sequence 512, App	797	36	42.4	716	4	US-10-127-830A-512	Sequence 512, App
725	36	42.4	716	4	US-10-131-830A-512	Sequence 512, App	798	36	42.4	716	4	US-10-127-832A-512	Sequence 512, App
726	36	42.4	716	4	US-10-131-837A-512	Sequence 512, App	799	36	42.4	716	4	US-10-127-833A-512	Sequence 512, App
727	36	42.4	716	4	US-10-137-872A-512	Sequence 512, App	800	36	42.4	716	4	US-10-127-834A-512	Sequence 512, App
728	36	42.4	716	4	US-10-147-500-512	Sequence 512, App	801	36	42.4	716	4	US-10-127-836A-512	Sequence 512, App
729	36	42.4	716	4	US-10-147-502-512	Sequence 512, App	802	36	42.4	716	4	US-10-127-841A-512	Sequence 512, App
730	36	42.4	716	4	US-10-147-515-512	Sequence 512, App	803	36	42.4	716	4	US-10-127-844A-512	Sequence 512, App
731	36	42.4	716	4	US-10-147-526-512	Sequence 512, App	804	36	42.4	716	4	US-10-128-687A-512	Sequence 512, App
732	36	42.4	716	4	US-10-147-527-512	Sequence 512, App	805	36	42.4	716	4	US-10-128-688A-512	Sequence 512, App
733	36	42.4	716	4	US-10-147-527-512	Sequence 512, App	806	36	42.4	716	4	US-10-128-689A-512	Sequence 512, App
734	36	42.4	716	4	US-10-121-041-512	Sequence 512, App	807	36	42.4	716	4	US-10-128-694A-512	Sequence 512, App
735	36	42.4	716	4	US-10-121-043-512	Sequence 512, App	808	36	42.4	716	4	US-10-128-694A-512	Sequence 512, App
736	36	42.4	716	4	US-10-121-047-512	Sequence 512, App	809	36	42.4	716	4	US-10-131-825A-512	Sequence 512, App
737	36	42.4	716	4	US-10-123-215-512	Sequence 512, App	810	36	42.4	716	4	US-10-230-417-512	Sequence 190, App
738	36	42.4	716	4	US-10-123-902-512	Sequence 512, App	811	36	42.4	716	4	US-10-219-003-190	Sequence 190, App
739	36	42.4	716	4	US-10-123-908-512	Sequence 512, App	812	36	42.4	716	4	US-10-219-075-190	Sequence 190, App
740	36	42.4	716	4	US-10-123-909-512	Sequence 512, App	813	36	42.4	716	4	US-10-219-464-190	Sequence 190, App
741	36	42.4	716	4	US-10-123-910-512	Sequence 512, App	814	36	42.4	716	4	US-10-219-466-190	Sequence 190, App
742	36	42.4	716	4	US-10-124-813-512	Sequence 512, App	815	36	42.4	716	4	US-10-219-479-190	Sequence 190, App
743	36	42.4	716	4	US-10-124-817-512	Sequence 512, App	816	36	42.4	716	4	US-10-219-481-190	Sequence 190, App
744	36	42.4	716	4	US-10-125-922-512	Sequence 512, App	817	36	42.4	716	4	US-10-230-260-190	Sequence 190, App
745	36	42.4	716	4	US-10-125-924-512	Sequence 512, App	818	36	42.4	716	4	US-10-232-231-190	Sequence 190, App
746	36	42.4	716	4	US-10-140-860-512	Sequence 512, App	819	36	42.4	716	4	US-10-232-233-190	Sequence 190, App
747	36	42.4	716	4	US-10-142-417-512	Sequence 512, App	820	36	42.4	716	4	US-10-131-815A-512	Sequence 512, App
748	36	42.4	716	4	US-10-147-519-512	Sequence 512, App	821	36	42.4	716	4	US-10-131-817A-512	Sequence 512, App
749	36	42.4	716	4	US-10-157-782-512	Sequence 512, App	822	36	42.4	716	4	US-10-131-821A-512	Sequence 512, App
750	36	42.4	716	4	US-10-152-395-512	Sequence 512, App	823	36	42.4	716	4	US-10-131-822A-512	Sequence 512, App
751	36	42.4	716	4	US-10-219-076-190	Sequence 190, App	824	36	42.4	716	4	US-10-131-828A-512	Sequence 512, App
752	36	42.4	716	4	US-10-230-434-190	Sequence 190, App	825	36	42.4	716	4	US-10-131-835A-512	Sequence 512, App
753	36	42.4	716	4	US-10-125-926A-512	Sequence 512, App	826	36	42.4	716	4	US-10-137-864A-512	Sequence 512, App
754	36	42.4	716	4	US-10-125-930A-512	Sequence 512, App	827	36	42.4	716	4	US-10-137-869A-512	Sequence 512, App
755	36	42.4	716	4	US-10-127-831A-512	Sequence 512, App	828	36	42.4	716	4	US-10-147-523-512	Sequence 512, App
756	36	42.4	716	4	US-10-127-837A-512	Sequence 512, App	829	36	42.4	716	4	US-10-158-785-512	Sequence 512, App
757	36	42.4	716	4	US-10-127-838B-512	Sequence 512, App	830	36	42.4	716	4	US-10-121-051-512	Sequence 512, App
												US-10-216-165-190	Sequence 190, App

831	Sequence 190, App	US-10-218-956-190	716	4	42.4	716	4	US-10-152-380-512	Sequence 512, App
832	Sequence 190, App	US-10-219-468-190	716	4	42.4	716	4	US-10-153-934-512	Sequence 512, App
833	Sequence 190, App	US-10-219-478-190	716	4	42.4	716	4	US-10-140-807-512	Sequence 512, App
834	Sequence 190, App	US-10-219-536-190	716	4	42.4	716	4	US-10-140-924-512	Sequence 512, App
835	Sequence 190, App	US-10-233-205-190	716	4	42.4	716	4	US-10-140-926-512	Sequence 512, App
836	Sequence 512, App	US-10-121-042-512	716	4	42.4	716	4	US-10-141-698-512	Sequence 512, App
837	Sequence 190, App	US-10-219-072-190	716	4	42.4	716	4	US-10-141-702-512	Sequence 512, App
838	Sequence 190, App	US-10-219-470-190	716	4	42.4	716	4	US-10-141-704-512	Sequence 512, App
839	Sequence 190, App	US-10-219-474-190	716	4	42.4	716	4	US-10-142-421-512	Sequence 512, App
840	Sequence 190, App	US-10-219-524-190	716	4	42.4	716	4	US-10-142-432-512	Sequence 512, App
841	Sequence 190, App	US-10-219-528-190	716	4	42.4	716	4	US-10-142-767-512	Sequence 512, App
842	Sequence 190, App	US-10-227-880-190	716	4	42.4	716	4	US-10-143-033-512	Sequence 512, App
843	Sequence 190, App	US-10-227-881-190	716	4	42.4	716	4	US-10-144-994-512	Sequence 512, App
844	Sequence 190, App	US-10-227-882-190	716	4	42.4	716	4	US-10-145-628-512	Sequence 512, App
845	Sequence 190, App	US-10-230-436-190	716	4	42.4	716	4	US-10-145-746-512	Sequence 512, App
846	Sequence 190, App	US-10-232-223-190	716	4	42.4	716	4	US-10-145-748-512	Sequence 512, App
847	Sequence 190, App	US-10-232-225-190	716	4	42.4	716	4	US-10-145-823-512	Sequence 512, App
848	Sequence 190, App	US-10-232-227-190	716	4	42.4	716	4	US-10-145-826-512	Sequence 512, App
849	Sequence 190, App	US-10-232-229-190	716	4	42.4	716	4	US-10-145-870-512	Sequence 512, App
850	Sequence 190, App	US-10-232-234-190	716	4	42.4	716	4	US-10-145-876-512	Sequence 512, App
851	Sequence 190, App	US-10-219-060-190	716	4	42.4	716	4	US-10-145-959-512	Sequence 512, App
852	Sequence 512, App	US-10-123-912-512	716	4	42.4	716	4	US-10-146-724-512	Sequence 512, App
853	Sequence 222, App	US-10-223-085-222	716	4	42.4	716	4	US-10-146-725-512	Sequence 512, App
854	Sequence 190, App	US-10-216-160-190	716	4	42.4	716	4	US-10-146-795-512	Sequence 512, App
855	Sequence 190, App	US-10-216-162-190	716	4	42.4	716	4	US-10-147-495-512	Sequence 512, App
856	Sequence 190, App	US-10-216-164-190	716	4	42.4	716	4	US-10-147-501-512	Sequence 512, App
857	Sequence 190, App	US-10-216-167-190	716	4	42.4	716	4	US-10-147-504-512	Sequence 512, App
858	Sequence 190, App	US-10-216-168-190	716	4	42.4	716	4	US-10-147-506-512	Sequence 512, App
859	Sequence 190, App	US-10-219-065-190	716	4	42.4	716	4	US-10-147-509-512	Sequence 512, App
860	Sequence 190, App	US-10-219-071-190	716	4	42.4	716	4	US-10-147-510-512	Sequence 512, App
861	Sequence 190, App	US-10-219-074-190	716	4	42.4	716	4	US-10-147-511-512	Sequence 512, App
862	Sequence 190, App	US-10-219-077-190	716	4	42.4	716	4	US-10-147-529-512	Sequence 512, App
863	Sequence 190, App	US-10-219-465-190	716	4	42.4	716	4	US-10-152-397-512	Sequence 512, App
864	Sequence 190, App	US-10-219-467-190	716	4	42.4	716	4	US-10-153-586-512	Sequence 512, App
865	Sequence 190, App	US-10-219-469-190	716	4	42.4	716	4	US-10-158-786-512	Sequence 512, App
866	Sequence 190, App	US-10-219-471-190	716	4	42.4	716	4	US-10-137-870-512	Sequence 512, App
867	Sequence 190, App	US-10-219-473-190	716	4	42.4	716	4	US-10-140-018-512	Sequence 512, App
868	Sequence 190, App	US-10-219-476-190	716	4	42.4	716	4	US-10-140-021-512	Sequence 512, App
869	Sequence 190, App	US-10-219-482-190	716	4	42.4	716	4	US-10-140-471-512	Sequence 512, App
870	Sequence 190, App	US-10-227-874-190	716	4	42.4	716	4	US-10-140-922-512	Sequence 512, App
871	Sequence 190, App	US-10-227-876-190	716	4	42.4	716	4	US-10-145-631-512	Sequence 512, App
872	Sequence 190, App	US-10-227-878-190	716	4	42.4	716	4	US-10-145-633-512	Sequence 512, App
873	Sequence 190, App	US-10-229-974-190	716	4	42.4	716	4	US-10-158-783-512	Sequence 512, App
874	Sequence 190, App	US-10-230-034-190	716	4	42.4	716	4	US-10-140-274-512	Sequence 512, App
875	Sequence 190, App	US-10-230-113-190	716	4	42.4	716	4	US-10-221-097-52	Sequence 52, Appl
876	Sequence 190, App	US-10-230-183-190	716	4	42.4	716	4	US-10-140-019-52	Sequence 512, App
877	Sequence 190, App	US-10-230-234-190	716	4	42.4	716	4	US-10-140-022-512	Sequence 512, App
878	Sequence 190, App	US-10-230-366-190	716	4	42.4	716	4	US-10-140-861-512	Sequence 512, App
879	Sequence 190, App	US-10-230-426-190	716	4	42.4	716	4	US-10-140-862-512	Sequence 512, App
880	Sequence 190, App	US-10-230-427-190	716	4	42.4	716	4	US-10-141-697-512	Sequence 512, App
881	Sequence 190, App	US-10-230-433-190	716	4	42.4	716	4	US-10-141-700-512	Sequence 512, App
882	Sequence 190, App	US-10-230-435-190	716	4	42.4	716	4	US-10-141-705-512	Sequence 512, App
883	Sequence 190, App	US-10-230-438-190	716	4	42.4	716	4	US-10-141-753-512	Sequence 512, App
884	Sequence 190, App	US-10-232-222-190	716	4	42.4	716	4	US-10-141-758-512	Sequence 512, App
885	Sequence 512, App	US-10-232-222-190	716	4	42.4	716	4	US-10-142-418-512	Sequence 512, App
886	Sequence 512, App	US-10-194-359-512	716	4	42.4	716	4	US-10-142-420-512	Sequence 512, App
887	Sequence 222, App	US-10-223-084-222	716	4	42.4	716	4	US-10-142-422-512	Sequence 512, App
888	Sequence 222, App	US-10-223-088-222	716	4	42.4	716	4	US-10-142-427-512	Sequence 512, App
889	Sequence 222, App	US-10-223-090-222	716	4	42.4	716	4	US-10-142-760-512	Sequence 512, App
890	Sequence 190, App	US-10-219-070-190	716	4	42.4	716	4	US-10-142-761-512	Sequence 512, App
891	Sequence 190, App	US-10-219-472-190	716	4	42.4	716	4	US-10-142-763-512	Sequence 512, App
892	Sequence 190, App	US-10-219-527-190	716	4	42.4	716	4	US-10-152-531-512	Sequence 512, App
893	Sequence 190, App	US-10-227-877-190	716	4	42.4	716	4	US-10-126-163-190	Sequence 190, App
894	Sequence 222, App	US-10-223-087-222	716	4	42.4	716	4	US-10-142-424-512	Sequence 512, App
895	Sequence 512, App	US-10-127-847A-512	716	4	42.4	716	4	US-10-142-761-512	Sequence 512, App
896	Sequence 222, App	US-10-223-083-222	716	4	42.4	716	4	US-10-142-763-512	Sequence 512, App
897	Sequence 190, App	US-10-216-166-190	716	4	42.4	716	4	US-10-142-765-512	Sequence 512, App
898	Sequence 190, App	US-10-218-612-190	716	4	42.4	716	4	US-10-142-887-512	Sequence 512, App
899	Sequence 222, App	US-10-223-089-222	716	4	42.4	716	4	US-10-142-888-512	Sequence 512, App
900	Sequence 512, App	US-10-137-866-512	716	4	42.4	716	4	US-10-143-034-512	Sequence 512, App
901	Sequence 512, App	US-10-146-726-512	716	4	42.4	716	4	US-10-143-116-512	Sequence 512, App
902	Sequence 512, App	US-10-146-727-512	716	4	42.4	716	4	US-10-144-957-512	Sequence 512, App
903	Sequence 512, App	US-10-146-788-512	716	4	42.4	716	4	US-10-144-992-512	Sequence 512, App

```
977 36 42.4 716 4 US-10-145-015-512 Sequence 512, App
978 36 42.4 716 4 US-10-145-090-512 Sequence 512, App
979 36 42.4 716 4 US-10-145-091-512 Sequence 512, App
980 36 42.4 716 4 US-10-145-629-512 Sequence 512, App
981 36 42.4 716 4 US-10-145-630-512 Sequence 512, App
982 36 42.4 716 4 US-10-145-747-512 Sequence 512, App
983 36 42.4 716 4 US-10-145-752-512 Sequence 512, App
984 36 42.4 716 4 US-10-145-754-512 Sequence 512, App
985 36 42.4 716 4 US-10-145-755-512 Sequence 512, App
986 36 42.4 716 4 US-10-145-818-512 Sequence 512, App
987 36 42.4 716 4 US-10-145-820-512 Sequence 512, App
988 36 42.4 716 4 US-10-145-872-512 Sequence 512, App
989 36 42.4 716 4 US-10-145-873-512 Sequence 512, App
990 36 42.4 716 4 US-10-147-481-512 Sequence 512, App
991 36 42.4 716 4 US-10-147-482-512 Sequence 512, App
992 36 42.4 716 4 US-10-147-503-512 Sequence 512, App
993 36 42.4 716 4 US-10-147-522-512 Sequence 512, App
994 36 42.4 716 4 US-10-152-401-512 Sequence 512, App
995 36 42.4 716 4 US-10-157-783-512 Sequence 512, App
996 36 42.4 716 4 US-10-158-792-512 Sequence 512, App
997 36 42.4 716 4 US-10-158-462-512 Sequence 512, App
998 36 42.4 716 4 US-10-143-035-512 Sequence 512, App
999 36 42.4 716 4 US-10-145-751-512 Sequence 512, App
1000 36 42.4 716 4 US-10-145-822-512 Sequence 512, App
```

## ALIGNMENTS

```
RESULT 1
US-10-383-930-28
; Sequence 28, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 28
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-28

Query Match 100.0%; Score 85; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 QWNGESEKPYDDHL 14
| | | | | | | | | |
Db 1 QWNGESEKPYDDHL 14
```

```
RESULT 2
US-10-797-821-28
; Sequence 28, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
```

```
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 28
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Surface Domain GTF Peptide
US-10-797-821-28
```

```
Query Match 100.0%; Score 85; DB 5; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.3e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 QWNGESEKPYDDHL 14
| | | | | | | | | |
Db 1 QWNGESEKPYDDHL 14
```

```
RESULT 3
US-10-797-302-4
; Sequence 4, Application US/10797302
; Publication No. US20050026271A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-019
; CURRENT APPLICATION NUMBER: US/10/797,302
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 4
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: MAC peptide
US-10-797-302-4
```

```
Query Match 100.0%; Score 85; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 QWNGESEKPYDDHL 14
| | | | | | | | | |
Db 2 QWNGESEKPYDDHL 15
```

RESULT 4

```
US-10-797-302-8
; Sequence 8, Application US/10797302
; Publication No. US20050026271A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-019
; CURRENT APPLICATION NUMBER: US/10/797,302
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Streptococcus downei
US-10-797-302-8

Query Match      100.0%; Score 85; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QWNGESEKPYDDHL 14
Db      2 QWNGESEKPYDDHL 15
      |||||
      |||||

RESULT 5
US-10-797-302-9
; Sequence 9, Application US/10797302
; Publication No. US20050026271A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-019
; CURRENT APPLICATION NUMBER: US/10/797,302
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-02
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 9
; LENGTH: 15
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-302-9

Query Match      100.0%; Score 85; DB 5; Length 15;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QWNGESEKPYDDHL 14
Db      2 QWNGESEKPYDDHL 15
      |||||
      |||||

US-10-797-821-28.rapbm
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match      100.0%; Score 85; DB 4; Length 1590;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QWNGESEKPYDDHL 14
Db      337 QWNGESEKPYDDHL 350
      |||||
      |||||

RESULT 6
US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match      100.0%; Score 85; DB 5; Length 1590;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 QWNGESEKPYDDHL 14
Db      337 QWNGESEKPYDDHL 350
      |||||
      |||||

RESULT 7
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match      100.0%; Score 85; DB 5; Length 1590;
Best Local Similarity 100.0%; Pred. No. 0.00017;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```





; ORGANISM: Streptococcus mutans  
US-10-797-821-35

Query Match 78.8%; Score 67; DB 5; Length 1375;  
Best Local Similarity 76.9%; Pred. No. 0.11;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 WNGSEKPYDDHL 14  
||:||||:||||  
Db 370 WNSDSEKPFDDHL 382

## RESULT 12

US-09-740-274-2  
; Sequence 2, Application US/09740274  
; Patent No. US20020031826A1  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 78.8%; Score 67; DB 3; Length 1475;  
Best Local Similarity 76.9%; Pred. No. 0.12;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 WNGSEKPYDDHL 14  
||:||||:||||  
Db 344 WNSDSEKPFDDHL 356

## RESULT 13

US-10-383-930-34  
; Sequence 34, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 34  
; LENGTH: 1475  
; TYPE: PRT

; ORGANISM: Streptococcus mutans  
US-10-383-930-34

Query Match 78.8%; Score 67; DB 4; Length 1475;  
Best Local Similarity 76.9%; Pred. No. 0.12;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 WNGSEKPYDDHL 14  
||:||||:||||  
Db 344 WNSDSEKPFDDHL 356

## RESULT 14

US-10-797-821-34  
; Sequence 34, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 34  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-34

Query Match 78.8%; Score 67; DB 5; Length 1475;  
Best Local Similarity 76.9%; Pred. No. 0.12;  
Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 WNGSEKPYDDHL 14  
||:||||:||||  
Db 344 WNSDSEKPFDDHL 356

## RESULT 15

US-10-797-302-5  
; Sequence 5, Application US/10797302  
; Publication No. US20050026271A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-019  
; CURRENT APPLICATION NUMBER: US/10/797,302  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07

; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-02  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 5  
; LENGTH: 15  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-302-5

Query Match 75.3%; Score 64; DB 5; Length 15;  
Best Local Similarity 69.2%; Pred. No. 0.0032;  
Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 WNGESEKPYDHL 14  
:|||||:|  
Db 3 WNSDSRPFDDHL 15

RESULT 16  
US-10-282-122A-78042  
; Sequence 78042, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 78042  
; LENGTH: 105  
; TYPE: PRT  
; ORGANISM: Yersinia pestis  
US-10-282-122A-78042

Query Match 56.5%; Score 48; DB 4; Length 105;  
Best Local Similarity 53.8%; Pred. No. 8.4;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QWNGSEKPYDDH 13  
:|||||:|  
Db 3 EWNGEYVSPYAEH 15

RESULT 17  
US-10-282-122A-68515  
; Sequence 68515, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09  
; PRIOR APPLICATION NUMBER: 60/269,308  
; PRIOR FILING DATE: 2001-02-16  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 78614  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 68515  
; LENGTH: 105  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-10-282-122A-68515

Query Match 55.3%; Score 47; DB 4; Length 105;  
Best Local Similarity 53.8%; Pred. No. 12;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 QWNGSEKPYDDH 13  
:|||||:|  
Db 3 EWNGEYVSPYAEH 15

RESULT 18  
US-10-484-218-22  
; Sequence 22, Application US/10484218  
; Publication No. US20050059633A1  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERITDINA HENDRIKA  
; TITLE OF INVENTION: GLUCANS AND GLUCANSUCRASES DERIVED FROM  
; TITLE OF INVENTION: LACTIC ACID BACTERIA  
; FILE REFERENCE: 2001-1316

; CURRENT APPLICATION NUMBER: US/10/484,218  
; CURRENT FILING DATE: 2004-01-20  
; PRIOR APPLICATION NUMBER: PCT/NL02/00495  
; PRIOR FILING DATE: 2002-07-22  
; PRIOR APPLICATION NUMBER: EP 01202752.0  
; PRIOR FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: EP 01202841.1  
; PRIOR FILING DATE: 2001-07-25  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 22  
; LENGTH: 1006  
; TYPE: PRT  
; ORGANISM: Lactobacillus fermentum  
US-10-484-218-22

Query Match 55.3%; Score 47; DB 5; Length 1006;  
Best Local Similarity 64.3%; Pred. No. 1.2e+02;  
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 QWNGSEKPYDDHL 14  
| | | | |  
Db 11 QWNKTSVDVNDHL 24

RESULT 19  
US-10-724-972A-6956  
; Sequence 6956, Application US/10724972A  
; Publication No. US20040147734A1  
; GENERAL INFORMATION:  
; APPLICANT: Doucette-Stamm, Lynn  
; APPLICANT: Bush, David  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
; FILE REFERENCE: BATH03-16  
; CURRENT APPLICATION NUMBER: US/10/724,972A  
; CURRENT FILING DATE: 2003-12-01  
; PRIOR APPLICATION NUMBER: 09/450,969  
; PRIOR FILING DATE: 1999-11-29  
; PRIOR APPLICATION NUMBER: 09/134,001  
; PRIOR FILING DATE: 1998-08-13  
; PRIOR APPLICATION NUMBER: 60/064,964  
; PRIOR FILING DATE: 1997-11-08  
; PRIOR APPLICATION NUMBER: 60/055,779  
; PRIOR FILING DATE: 1997-08-14  
; NUMBER OF SEQ ID NOS: 7544  
; SEQ ID NO 6956  
; LENGTH: 504  
; TYPE: PRT  
; ORGANISM: S.epidermidis  
US-10-724-972A-6956

Query Match 54.1%; Score 46; DB 4; Length 504;  
Best Local Similarity 54.5%; Pred. No. 88;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 2 WNGSEKPYDD 12  
| | | | |  
Db 420 WRGTDNRPYED 430

RESULT 20  
US-10-437-963-153309  
; Sequence 153309, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad

; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 153309  
; LENGTH: 864  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_53277C.1.pep  
US-10-437-963-153309

Query Match 54.1%; Score 46; DB 4; Length 864;  
Best Local Similarity 63.6%; Pred. No. 1.5e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WNGSEKPYDD 12  
| | | | |  
Db 706 WGGEGGKPWDD 716

RESULT 21  
US-10-424-599-151656  
; Sequence 151656, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 151656  
; LENGTH: 59  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_107969C.1.pep  
US-10-424-599-151656

Query Match 52.9%; Score 45; DB 4; Length 59;  
Best Local Similarity 70.0%; Pred. No. 14;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ESEKPYDDHL 14  
| | | | |  
Db 24 EASQPYDDHL 33

RESULT 22  
US-10-282-122A-77550  
; Sequence 77550, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haeelbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.

```
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77550
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Vibrio cholerae
US-10-282-122A-77550

Query Match      52.9%; Score 45; DB 4; Length 105;
Best Local Similarity 58.3%; Pred. No. 25;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      2 WNGESEKPYDDH 13
Db      4 WNGEYISFYAEH 15

RESULT 23
US-10-424-599-262066
; Sequence 262066, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 262066
; LENGTH: 178
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_78668C.1.pep
US-10-424-599-262066

Query Match      52.9%; Score 45; DB 4; Length 178;
Best Local Similarity 57.1%; Pred. No. 44;
Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      1 QWNGESEKPYDDH 14
Db      84 QYLGNSSEPMNDH 97
```

```
RESULT 24
US-10-425-115-345898
; Sequence 345898, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 345898
; LENGTH: 124
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_78627C.1.pep
US-10-425-115-345898

Query Match      51.8%; Score 44; DB 4; Length 124;
Best Local Similarity 38.5%; Pred. No. 43;
Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY      2 WNGESEKPYDDH 14
Db      2 WHADDERPYQTHI 14

RESULT 25
US-10-156-761-9453
; Sequence 9453, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9453
; LENGTH: 175
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9453

Query Match      51.8%; Score 44; DB 4; Length 175;
Best Local Similarity 53.8%; Pred. No. 62;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      2 WNGESEKPYDDH 14
Db      19 WPGKTVTEYDDH 31

RESULT 26
US-10-425-115-210862
; Sequence 210862, Application US/10425115
; Publication No. US20040214272A1
```

```
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 210862
; LENGTH: 425
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(425)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_123901C.1.pep
US-10-425-115-210862

Query Match          51.8%; Score 44; DB 4; Length 425;
Best Local Similarity 53.8%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      2  WNGSEKPYDDHL 14
Db      322 WNGEENPHGHGL 334

RESULT 27
US-10-425-114-65670
; Sequence 65670, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65670
; LENGTH: 480
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700335575_FLI.pep
US-10-425-114-65670

Query Match          51.8%; Score 44; DB 4; Length 480;
Best Local Similarity 53.8%; Pred. No. 1.7e+02;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      2  WNGSEKPYDDHL 14
Db      377 WNGEENPHGHGL 389

RESULT 28
US-10-190-115-28
; Sequence 28, Application US/10190115
; Publication No. US20030207394A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, John P. II
; APPLICANT: Boldog, Ferenc L.
```

```
; APPLICANT: Burgess, Catherine E.
; APPLICANT: Casman, Stacie J.
; APPLICANT: Grosse, William M.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Ji, Weizhen
; APPLICANT: Lepley, Denise M.
; APPLICANT: Liu, Xiaohong
; APPLICANT: Mezik, Amanda J.
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Patturajan, Meera
; APPLICANT: Rastelli, Luca
; APPLICANT: Shen, Lei
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Szekeres, Edward S. Jr.
; APPLICANT: Taupier, Raymond J. Jr.
; APPLICANT: Tchernev, Velizar T.
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Voss, Edward Z.
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-050 CIP
; CURRENT APPLICATION NUMBER: US/10/190,115
; CURRENT FILING DATE: 2003-02-10
; PRIOR APPLICATION NUMBER: 60/303,168
; PRIOR FILING DATE: 2001-07-05
; PRIOR APPLICATION NUMBER: 60/368,996
; PRIOR FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: 60/386,816
; PRIOR FILING DATE: 2002-06-07
; PRIOR APPLICATION NUMBER: 60/215,854
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,856
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,902
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/216,585,
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,586
; PRIOR FILING DATE: 2001-07-07
; PRIOR APPLICATION NUMBER: 60/216,722
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/218,622
; PRIOR FILING DATE: 2000-07-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 136
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 28
; LENGTH: 2447
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-190-115-28

Query Match          51.8%; Score 44; DB 4; Length 2447;
Best Local Similarity 46.2%; Pred. No. 9.3e+02;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY      2  WNGSEKPYDDHL 14
Db      2195 WGTTRKPYSDHW 2207

RESULT 29
US-10-369-072-28
; Sequence 28, Application US/10369072
; Publication No. US20040014081A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook II, John P
; APPLICANT: Spaderna, Stephen K
; APPLICANT: Tchernev, Velizar
; APPLICANT: Liu, Xiaohong
; APPLICANT: Shenoy, Suresh
```

```
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zerhusen, Bryan
; APPLICANT: Patturajan, Meera
; APPLICANT: Taupier, Raymond T
; APPLICANT: Rastelli, Luca
; APPLICANT: Grosse, William M
; APPLICANT: Szerkeres, Edward S
; APPLICANT: Lepley, Denise M
; APPLICANT: Shen, Lei
; APPLICANT: Burgess, Catherine E
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigaru, Muralidhara
; TITLE OF INVENTION: No. US20040014081A1e1 Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-050 CON2
; CURRENT APPLICATION NUMBER: US/10/369,072
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: 10/1174,372
; PRIOR FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: 09/898,994
; PRIOR FILING DATE: 2001-07-03
; PRIOR APPLICATION NUMBER: 60/215,854
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,856
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/215,902
; PRIOR FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/216,585
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,586
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/216,722
; PRIOR FILING DATE: 2000-07-07
; PRIOR APPLICATION NUMBER: 60/218,622
; PRIOR FILING DATE: 2000-07-17
; PRIOR APPLICATION NUMBER: 60/218,992
; PRIOR FILING DATE: 2000-07-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 28
; LENGTH: 2447
; TYPE: PRT
; ORGANISM: Takifugu rubripes
US-10-369-072-28
```

```
Query Match 51.8%; Score 44; DB 4; Length 2447;
Best Local Similarity 46.2%; Pred. No. 9.3e+02;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 2 WNGESEKPYDDHL 14
Db 2195 WVGTTKHPYSDHM 2207
```

```
RESULT 30
US-10-437-963-127860
; Sequence 127860, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
```

```
; SEQ ID NO 127860
; LENGTH: 114
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_3026C.1.pep
US-10-437-963-127860
```

```
Query Match 50.6%; Score 43; DB 4; Length 114;
Best Local Similarity 54.5%; Pred. No. 58;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 WNGESEKPYDD 12
Db 70 WGGEGGRPWDD 80
```

```
RESULT 31
US-10-437-963-127805
; Sequence 127805, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 127805
; LENGTH: 142
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_3021C.1.pep
US-10-437-963-127805
```

```
Query Match 50.6%; Score 43; DB 4; Length 142;
Best Local Similarity 54.5%; Pred. No. 72;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

```
QY 2 WNGESEKPYDD 12
Db 100 WGGEGGRPWDD 110
```

```
RESULT 32
US-10-437-963-195318
; Sequence 195318, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
```

```
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_91279C.1.pep
US-10-437-963-195318

Query Match      50.6%; Score 43; DB 4; Length 422;
Best Local Similarity 50.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QWGESEKPYDD 12
Db 127 KWEALDSPYDD 138

RESULT 33
US-10-437-963-170027
; Sequence 170027, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 170027
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)...(545)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_68391C.1.pep
US-10-437-963-170027

Query Match      50.6%; Score 43; DB 4; Length 545;
Best Local Similarity 53.8%; Pred. No. 2.9e+02;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 2 WNGESEKPYDHL 14
Db 287 WNGEGPNPLGHL 299

RESULT 34
US-10-437-963-127893
; Sequence 127893, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
```

```
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 127893
; LENGTH: 700
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_3029C.1.pep
US-10-437-963-127893

Query Match      50.6%; Score 43; DB 4; Length 700;
Best Local Similarity 54.5%; Pred. No. 3.7e+02;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WNGESEKPYDD 12
Db 547 WGEGGRPWDD 557

RESULT 35
US-11-097-143-29574
; Sequence 29574, Application US/11097143
; Publication No. US20050208558A1
; GENERAL INFORMATION:
; APPLICANT: Venter, J. Craig
; APPLICANT: et al.
; TITLE OF INVENTION: DETECTION KIT, SUCH AS NUCLEIC ACID
; TITLE OF INVENTION: ARRAYS, FOR DETECTING EXPRESSION OF 10,000 OR MORE
; TITLE OF INVENTION: DROSOPHILA GENES.
; FILE REFERENCE: CL000728
; CURRENT APPLICATION NUMBER: US/11/097,143
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: 60/157,832
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: 60/160,191
; PRIOR FILING DATE: 1999-10-19
; PRIOR APPLICATION NUMBER: 60/161,932
; PRIOR FILING DATE: 1999-10-28
; PRIOR APPLICATION NUMBER: 60/164,769
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/173,383
; PRIOR FILING DATE: 1999-12-28
; PRIOR APPLICATION NUMBER: 60/175,693
; PRIOR FILING DATE: 2000-01-12
; PRIOR APPLICATION NUMBER: 60/184,831
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: 60/191,637
; PRIOR FILING DATE: 2000-03-23
; NUMBER OF SEQ ID NOS: 43008
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 29574
; LENGTH: 268
; TYPE: PRT
; ORGANISM: DROSOPHILA
US-11-097-143-29574

Query Match      50.0%; Score 42.5; DB 6; Length 268;
Best Local Similarity 53.3%; Pred. No. 1.7e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

Qy 2 WNGESEKPYD---DH 13
Db 251 WSGGSSSPYDNGNDH 265

RESULT 36
US-10-282-122A-43094
; Sequence 43094, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
```

```

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 43094
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-282-122A-43094

Query Match 49.4%; Score 42; DB 4; Length 105;
Best Local Similarity 46.2%; Pred. No. 76;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QWNGSEKPYDDH 13
:|:|:|:|:|
Db 3 EWSGEYISPYAEH 15

RESULT 37
US-10-282-122A-56312
; Sequence 56312, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078

```

```

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 56312
; LENGTH: 105
; TYPE: PRT
; ORGANISM: Enterobacter cloacae
US-10-282-122A-56312

Query Match 49.4%; Score 42; DB 4; Length 105;
Best Local Similarity 46.2%; Pred. No. 76;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QWNGSEKPYDDH 13
:|:|:|:|:|
Db 3 EWSGEYISPYAEH 15

RESULT 38
US-10-282-122A-59859
; Sequence 59859, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/257,931

```





; APPLICANT: Hyseq, Inc  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
 ; FILE REFERENCE: 790CIP3/US  
 ; CURRENT APPLICATION NUMBER: US/10/450,763  
 ; CURRENT FILING DATE: 2003-06-11  
 ; PRIOR APPLICATION NUMBER: PCT/US01/08631  
 ; PRIOR FILING DATE: 2001-03-30  
 ; PRIOR APPLICATION NUMBER: 09/540,217  
 ; PRIOR FILING DATE: 2000-03-31  
 ; PRIOR APPLICATION NUMBER: 09/649,167  
 ; PRIOR FILING DATE: 2000-08-23  
 ; NUMBER OF SEQ ID NOS: 60736  
 ; SOFTWARE: Custom  
 ; SEQ ID NO 34912  
 ; LENGTH: 188  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)..(188)  
 ; OTHER INFORMATION: Xaa = X or \* as defined in Table 2  
 US-10-450-763-34912

Query Match 49.4%; Score 42; DB 5; Length 188;  
 Best Local Similarity 57.1%; Pred. No. 1.4e+02;  
 Matches 8; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

Qy 1 QWNGSEKPYDDHL 14  
 ||||| : |||||  
 Db 59 QWNGHN--PVQDHL 70

## RESULT 42

US-10-450-763-60378  
 ; Sequence 60378, Application US/10450763  
 ; Publication No. US20050196754A1  
 ; GENERAL INFORMATION:

; APPLICANT: Hyseq, Inc  
 ; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES  
 ; FILE REFERENCE: 790CIP3/US  
 ; CURRENT APPLICATION NUMBER: US/10/450,763  
 ; CURRENT FILING DATE: 2003-06-11  
 ; PRIOR APPLICATION NUMBER: PCT/US01/08631  
 ; PRIOR FILING DATE: 2001-03-30  
 ; PRIOR APPLICATION NUMBER: 09/540,217  
 ; PRIOR FILING DATE: 2000-03-31  
 ; PRIOR APPLICATION NUMBER: 09/649,167  
 ; PRIOR FILING DATE: 2000-08-23  
 ; NUMBER OF SEQ ID NOS: 60736  
 ; SOFTWARE: Custom  
 ; SEQ ID NO 60378  
 ; LENGTH: 299  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: DOMAIN  
 ; LOCATION: (196)..(225)  
 ; OTHER INFORMATION: Met Apo-repressor MetJ, domain identified by eMATRIX,  
 ; OTHER INFORMATION: accession number PF01340A, p-value=1.375e-37, raw score of 7.96  
 ; FEATURE:  
 ; NAME/KEY: DOMAIN  
 ; LOCATION: (196)..(299)  
 ; OTHER INFORMATION: Met Apo-repressor, MetJ domain identified by Pfam, accession  
 ; OTHER INFORMATION: name MetJ, E-value=7.8e-88, Pfam score of 305.2  
 US-10-450-763-60378

Query Match 49.4%; Score 42; DB 5; Length 299;  
 Best Local Similarity 46.2%; Pred. No. 2.2e+02;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 QWNGSEKPYDDH 13  
 :||| :|||  
 Db 197 EWSGEVISPYAEH 209

## RESULT 43

US-09-815-242-13466  
 ; Sequence 13466, Application US/09815242  
 ; Patent No. US20020061569A1  
 ; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari L.  
 ; APPLICANT: Zyskind, Judith W.  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John D.  
 ; APPLICANT: Carr, Grant J.  
 ; APPLICANT: Yamamoto, Robert T.  
 ; APPLICANT: Xu, H. Howard  
 ; TITLE OF INVENTION: Identification of Essential Genes in  
 ; TITLE OF INVENTION: Prokaryotes  
 ; FILE REFERENCE: ELITRA.011A  
 ; CURRENT APPLICATION NUMBER: US/09/815,242  
 ; CURRENT FILING DATE: 2001-03-21  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21  
 ; PRIOR APPLICATION NUMBER: 60/206,848  
 ; PRIOR FILING DATE: 2000-05-23  
 ; PRIOR APPLICATION NUMBER: 60/207,727  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 60/242,578  
 ; PRIOR FILING DATE: 2000-10-23  
 ; PRIOR APPLICATION NUMBER: 60/253,625  
 ; PRIOR FILING DATE: 2000-11-27  
 ; PRIOR APPLICATION NUMBER: 60/257,931  
 ; PRIOR FILING DATE: 2000-12-22  
 ; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; NUMBER OF SEQ ID NOS: 14110  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 13466  
 ; LENGTH: 424  
 ; TYPE: PRT  
 ; ORGANISM: Streptococcus pneumoniae  
 US-09-815-242-13466

Query Match 49.4%; Score 42; DB 3; Length 424;  
 Best Local Similarity 53.8%; Pred. No. 3.2e+02;  
 Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 WNGSEKPYDDHL 14  
 ||||| :||| :|||  
 Db 5 WNGFSKSYQERL 17

## RESULT 44

US-10-282-122A-74130  
 ; Sequence 74130, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:

; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078

```
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 74130
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-10-282-122A-74130

Query Match          49.4%; Score 42; DB 4; Length 424;
Best Local Similarity 53.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy      2 WNGSEKPYDDHL 14
      |||||:|:|:|
Db      5 WNGFSKKSQYERL 17

RESULT 45
US-10-041-018-171
; Sequence 171, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 171
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-10-041-018-171

Query Match          49.4%; Score 42; DB 4; Length 424;
Best Local Similarity 53.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy      2 WNGSEKPYDDHL 14
      |||||:|:|:|
Db      5 WNGFSKKSQYERL 17

RESULT 46
US-10-041-018-270
; Sequence 270, Application US/10041018
; Publication No. US20040072323A1
; GENERAL INFORMATION:
; APPLICANT: Matsuda, Seiichi P.T.
```

```
; APPLICANT: Hart, Elizabeth A.
; TITLE OF INVENTION: Diterpene-Producing Unicellular Organism
; FILE REFERENCE: P02080US1/10025547
; CURRENT APPLICATION NUMBER: US/10/041,018
; CURRENT FILING DATE: 2002-01-07
; PRIOR APPLICATION NUMBER: US 60/259880
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 413
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 270
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; US-10-041-018-270

Query Match          49.4%; Score 42; DB 4; Length 424;
Best Local Similarity 53.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy      2 WNGSEKPYDDHL 14
      |||||:|:|:|
Db      5 WNGFSKKSQYERL 17

RESULT 47
US-10-472-928-3562
; Sequence 3562, Application US/10472928
; Publication No. US2005020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SPA
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: SeqWin99, version 1.03
; SEQ ID NO 3562
; LENGTH: 424
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: 3-hydroxy-3-methylglutaryl-CoA reductase
; OTHER INFORMATION: Cellular location: membrane
; OTHER INFORMATION: Similar to strain R6 sequence 15903612 (O.E+01)
; US-10-472-928-3562

Query Match          49.4%; Score 42; DB 5; Length 424;
Best Local Similarity 53.8%; Pred. No. 3.2e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy      2 WNGSEKPYDDHL 14
      |||||:|:|:|
Db      5 WNGFSKKSQYERL 17

RESULT 48
US-10-617-320-3690
; Sequence 3690, Application US/10617320
; Publication No. US2005013640A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
```

; COUNTRY: USA  
; ZIP: 02354  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: CD/ROM ISO9660  
; COMPUTER: <Unknown>  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: <Unknown>  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/10/617,320  
; FILING DATE: 10-Jul-2003  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/09/107,433  
; FILING DATE: 30-Jun-1998  
; APPLICATION NUMBER: 60/ 085131  
; FILING DATE: May 12, 1998  
; APPLICATION NUMBER: 60/051553  
; FILING DATE: July 2, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-011  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277  
; INFORMATION FOR SEQ ID NO: 3690:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 426 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: YES  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus pneumoniae  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (B) LOCATION 1...426  
; SEQUENCE DESCRIPTION: SEQ ID NO: 3690:  
US-10-617-320-3690

Query Match 49.4%; Score 42; DB 5; Length 426;  
Best Local Similarity 53.8%; Pred. No. 3.2e+02;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 WNGESEKPYDHL 14  
Db 7 WNGFSKASYOERL 19

## RESULT 49

US-10-425-115-240837  
; Sequence 240837, Application US/10425115  
; Publication No. US20040214272A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53222)B  
; CURRENT APPLICATION NUMBER: US/10/425,115  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 369326  
; SEQ ID NO 240837  
; LENGTH: 492  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)...(492)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: MRT4577\_151221C.1.pep

## US-10-425-115-240837

Query Match 49.4%; Score 42; DB 4; Length 492;  
Best Local Similarity 70.0%; Pred. No. 3.7e+02;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 2 WNGESEKPYD 11  
Db 202 WRKPEKPYD 211

## RESULT 50

US-10-437-963-201769  
; Sequence 201769, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO 201769  
; LENGTH: 785  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(785)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_97111C.1.pep  
US-10-437-963-201769

Query Match 49.4%; Score 42; DB 4; Length 785;  
Best Local Similarity 50.0%; Pred. No. 6e+02;  
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 1 QWNGESEKPYDD 12  
Db 614 EWNEECQKDFDD 625

Search completed: February 10, 2006, 23:50:36  
Job time : 204 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:47:29 ; Search time 17 Seconds  
(without alignments)  
10.807 Million cell updates/sec

Title: US-10-797-821-28

Perfect score: 85

Sequence: 1 QWNGSEKPYDDHL 14

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 1000 summaries

Database : Published Applications AA New:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	52.9	179	7	US-11-098-686-10166
2	40	47.1	319	6	US-10-793-626-2760
3	39	45.9	309	6	US-10-793-626-356
4	38.5	45.3	603	7	US-11-024-959-484
5	37.5	44.1	332	7	US-11-129-143-53
6	37.5	44.1	488	7	US-11-024-959-474
7	37	43.5	168	6	US-10-821-234-1479
8	37	43.5	178	6	US-10-821-234-1479
9	37	43.5	2333	6	US-10-453-372-170
10	37	43.5	2662	6	US-10-453-372-114
11	37	43.5	2724	6	US-10-453-372-148
12	37	43.5	2733	6	US-10-453-372-136
13	37	43.5	2733	6	US-10-453-372-142
14	37	43.5	2733	6	US-10-453-372-146
15	37	43.5	2733	6	US-10-453-372-150
16	37	43.5	2733	6	US-10-453-372-154
17	37	43.5	2759	6	US-10-453-372-168
18	37	43.5	2765	6	US-10-453-372-116
19	36.5	42.9	293	7	US-11-006-031-82
20	36.5	42.9	293	7	US-11-194-110-82
21	36.5	42.9	296	7	US-11-194-110-48
22	36	42.4	248	7	US-11-052-554A-70
23	36	42.4	248	7	US-11-052-554A-342
24	36	42.4	317	7	US-11-129-143-85
25	36	42.4	716	6	US-10-131-826A-512

Sequence 52, Appli	716	7	US-11-147-047-52
Sequence 2, Appli	716	7	US-11-142-867-2
Sequence 348, App	799	7	US-11-074-176-348
Sequence 172, App	805	7	US-11-074-176-172
Sequence 679, App	825	6	US-10-995-561-679
Sequence 469, App	825	6	US-11-124-367A-469
Sequence 1179, Ap	829	6	US-10-821-234-1179
Sequence 5, Appli	852	7	US-11-104-523A-5
Sequence 54, Appli	1045	7	US-11-113-424-54
Sequence 388, App	1047	7	US-11-124-367A-388
Sequence 386, App	1058	7	US-11-124-367A-386
Sequence 387, App	1062	7	US-11-124-367A-387
Sequence 1097, Ap	1094	6	US-10-821-234-1097
Sequence 4, Appli	2376	7	US-11-096-051-4
Sequence 2, Appli	2715	7	US-11-096-051-2
Sequence 51, Appli	2715	7	US-11-113-424-51
Sequence 10, Appli	2721	7	US-11-096-051-10
Sequence 8, Appli	2725	7	US-11-096-051-8
Sequence 371, App	523	7	US-11-024-959-371
Sequence 2762, Ap	82	6	US-10-467-657-2762
Sequence 1513, Ap	244	7	US-11-054-515-1513
Sequence 215, App	305	7	US-11-156-084-215
Sequence 10238, A	350	7	US-11-098-686-10238
Sequence 32, Appli	350	6	US-10-063-703-32
Sequence 32, Appli	445	7	US-11-102-240-32
Sequence 28, Appli	445	6	US-10-508-263-28
Sequence 22, Appli	481	6	US-10-508-263-22
Sequence 20, Appli	485	6	US-10-508-263-20
Sequence 12, Appli	495	6	US-11-113-424-12
Sequence 47, Appli	668	7	US-11-113-424-47
Sequence 5, Appli	689	7	US-11-113-424-5
Sequence 99, Appli	700	6	US-10-467-657-99
Sequence 398, App	828	6	US-11-124-367A-398
Sequence 396, App	908	7	US-11-124-367A-396
Sequence 394, App	1308	7	US-11-124-367A-394
Sequence 395, App	1332	7	US-11-124-367A-395
Sequence 397, App	1413	7	US-11-124-367A-397
Sequence 14, Appli	1452	7	US-11-113-424-14
Sequence 11096, A	2769	7	US-11-098-686-11096
Sequence 234, App	108	7	US-10-467-657-294
Sequence 5944, Ap	124	6	US-10-467-657-5944
Sequence 7, Appli	125	7	US-11-207-721-7
Sequence 6, Appli	215	7	US-11-218-821-6
Sequence 216, App	305	7	US-11-156-084-216
Sequence 1388, Ap	317	6	US-10-821-234-1388
Sequence 218, App	342	6	US-10-689-742-218
Sequence 69, Appli	365	6	US-10-770-726-69
Sequence 10835, A	513	7	US-11-098-686-10835
Sequence 2, Appli	562	7	US-11-207-721-2
Sequence 40, Appli	621	6	US-10-131-826A-40
Sequence 10148, A	789	7	US-11-098-686-10148
Sequence 4, Appli	1061	7	US-11-121-438-4
Sequence 5870, Ap	206	6	US-10-467-657-5870
Sequence 10753, A	457	7	US-11-098-686-10753
Sequence 1, Appli	1295	7	US-11-091-528-1
Sequence 4606, Ap	1403	7	US-11-091-528-2
Sequence 2612, Ap	83	6	US-10-467-657-4606
Sequence 101, Appli	174	6	US-10-793-626-2612
Sequence 10387, A	271	7	US-11-179-977-10
Sequence 285, App	295	7	US-11-098-686-10387
Sequence 61, Appli	339	7	US-11-124-367A-265
Sequence 267, App	344	6	US-10-922-232B-61
Sequence 5, Appli	352	7	US-11-124-367A-267
Sequence 6, Appli	372	7	US-11-143-986-5
Sequence 2, Appli	372	7	US-11-143-986-6
Sequence 233, App	386	7	US-11-143-986-2
Sequence 1, Appli	386	7	US-11-143-986-3
Sequence 604, App	397	7	US-11-022-562-223
Sequence 286, App	421	6	US-10-858-730-1
Sequence 11, Appli	522	6	US-10-793-626-604
Sequence 286, App	524	7	US-11-124-368A-286
Sequence 97, Appli	531	7	US-11-076-074-11
Sequence 97, Appli	542	7	US-11-019-711-97

99	33	38.8	556	7	US-11-124-368A-287	Sequence 287, App	172	31	36.5	223	7	US-11-124-367A-317	Sequence 317, App
100	33	38.8	643	7	US-11-054-281-113	Sequence 113, App	173	31	36.5	229	7	US-11-124-925-2	Sequence 2, Appli
101	33	38.8	685	7	US-11-098-686-10574	Sequence 10574, A	174	31	36.5	240	7	US-11-134-241-33	Sequence 33, Appl
102	33	38.8	689	7	US-11-024-559-510	Sequence 510, App	175	31	36.5	257	6	US-10-527-500-67	Sequence 67, Appl
103	33	38.8	857	7	US-11-052-554A-218	Sequence 218, App	176	31	36.5	311	6	US-10-131-826A-32	Sequence 32, Appl
104	33	38.8	907	7	US-11-124-368A-288	Sequence 288, App	177	31	36.5	311	6	US-10-512-214-16	Sequence 16, Appl
105	33	38.8	1480	7	US-11-076-074-101	Sequence 10, Appl	178	31	36.5	321	6	US-11-134-241-16	Sequence 35, Appl
106	33	38.8	1823	6	US-10-995-561-988	Sequence 988, App	179	31	36.5	326	7	US-11-229-371-75	Sequence 75, Appl
107	33	38.8	2102	6	US-10-995-561-990	Sequence 990, App	180	31	36.5	347	7	US-11-098-686-10337	Sequence 10337, A
108	33	38.8	2108	6	US-10-995-561-989	Sequence 989, App	181	31	36.5	366	7	US-11-075-351-38	Sequence 38, Appl
109	33	38.8	2157	6	US-10-995-561-991	Sequence 991, App	182	31	36.5	366	7	US-11-134-241-39	Sequence 39, Appl
110	33	38.8	6738	6	US-10-922-232B-56	Sequence 56, Appl	183	31	36.5	368	7	US-11-214-796-4	Sequence 4, Appli
111	32.5	38.2	152	7	US-11-052-554A-19	Sequence 19, Appl	184	31	36.5	374	7	US-11-075-351-42	Sequence 42, Appl
112	32.5	38.2	294	6	US-10-055-877-128	Sequence 128, App	185	31	36.5	384	6	US-10-467-657-2024	Sequence 2024, Ap
113	32.5	38.2	307	6	US-10-055-877-263	Sequence 263, App	186	31	36.5	385	7	US-11-052-554A-132	Sequence 132, App
114	32.5	38.2	333	6	US-10-873-528-43	Sequence 43, Appl	187	31	36.5	413	7	US-11-032-951-12	Sequence 12, Appl
115	32.5	38.2	349	6	US-10-131-826A-170	Sequence 170, App	188	31	36.5	414	6	US-10-631-558-2	Sequence 2, Appli
116	32.5	38.2	368	7	US-11-053-185-12	Sequence 12, Appl	189	31	36.5	448	6	US-10-467-657-1096	Sequence 1096, Ap
117	32.5	38.2	491	7	US-11-053-185-22	Sequence 22, Appl	190	31	36.5	451	6	US-10-467-657-366	Sequence 366, App
118	32	37.6	209	7	US-11-098-686-11039	Sequence 11039, A	191	31	36.5	508	7	US-11-082-389-178	Sequence 178, App
119	32	37.6	246	7	US-11-054-515-1702	Sequence 1702, Ap	192	31	36.5	543	6	US-10-858-730-221	Sequence 221, App
120	32	37.6	254	6	US-10-507-786-3	Sequence 3, Appli	193	31	36.5	552	7	US-11-219-995-9	Sequence 9, Appli
121	32	37.6	274	6	US-10-507-786-2	Sequence 2, Appli	194	31	36.5	575	6	US-10-995-805-2	Sequence 2, Appli
122	32	37.6	316	7	US-11-241-347-12	Sequence 12, App	195	31	36.5	603	7	US-11-134-241-43	Sequence 43, Appl
123	32	37.6	332	6	US-10-454-437-308	Sequence 308, App	196	31	36.5	619	6	US-10-517-939-352	Sequence 352, App
124	32	37.6	348	7	US-11-098-686-8740	Sequence 8740, Ap	197	31	36.5	631	7	US-11-134-241-41	Sequence 41, Appl
125	32	37.6	348	7	US-11-098-686-11203	Sequence 11203, A	198	31	36.5	658	7	US-11-098-686-11018	Sequence 11018, A
126	32	37.6	351	7	US-11-241-347-11	Sequence 11, Appl	199	31	36.5	688	7	US-11-098-686-10191	Sequence 10191, A
127	32	37.6	359	7	US-11-098-686-10361	Sequence 10361, A	200	31	36.5	696	7	US-11-242-243-2	Sequence 2, Appli
128	32	37.6	361	7	US-11-040-595-2	Sequence 2, Appli	201	31	36.5	708	7	US-11-174-150-25	Sequence 25, Appl
129	32	37.6	365	6	US-10-454-437-394	Sequence 394, App	202	31	36.5	736	7	US-11-174-150-26	Sequence 26, Appl
130	32	37.6	432	6	US-10-836-953-1	Sequence 1, Appli	203	31	36.5	736	7	US-11-078-189-9	Sequence 9, Appli
131	32	37.6	442	6	US-10-793-626-1000	Sequence 1000, Ap	204	31	36.5	737	6	US-10-878-556A-25	Sequence 25, Appl
132	32	37.6	447	7	US-11-169-041-142	Sequence 142, App	205	31	36.5	739	7	US-11-078-189B-12	Sequence 12, Appl
133	32	37.6	475	6	US-10-873-528-7	Sequence 7, Appli	206	31	36.5	767	6	US-10-467-962B-91	Sequence 91, Appl
134	32	37.6	486	6	US-10-467-657-950	Sequence 950, App	207	31	36.5	832	6	US-10-512-109-29	Sequence 29, Appl
135	32	37.6	512	7	US-11-112-824-22	Sequence 22, Appl	208	31	36.5	869	7	US-11-113-751-2	Sequence 2, Appli
136	32	37.6	512	7	US-11-112-824-23	Sequence 23, Appl	209	31	36.5	872	7	US-11-077-550-145	Sequence 145, App
137	32	37.6	512	7	US-11-112-824-24	Sequence 24, Appl	210	31	36.5	879	7	US-11-077-550-143	Sequence 143, App
138	32	37.6	512	7	US-11-112-824-25	Sequence 25, Appl	211	31	36.5	887	7	US-11-077-550-147	Sequence 147, App
139	32	37.6	512	7	US-11-112-824-26	Sequence 26, Appl	212	31	36.5	1033	7	US-11-037-243-75	Sequence 75, Appl
140	32	37.6	512	7	US-11-112-824-27	Sequence 27, Appl	213	31	36.5	1100	7	US-11-113-751-4	Sequence 4, Appli
141	32	37.6	513	7	US-11-112-824-1	Sequence 1, Appli	214	31	36.5	1115	7	US-11-113-751-14	Sequence 14, Appl
142	32	37.6	548	6	US-10-467-657-4672	Sequence 4672, Ap	215	31	36.5	1169	7	US-11-098-686-10502	Sequence 10502, A
143	32	37.6	549	6	US-10-467-657-4612	Sequence 4612, Ap	216	31	36.5	1315	7	US-11-077-550-141	Sequence 141, App
144	32	37.6	549	6	US-10-467-657-7190	Sequence 7190, Ap	217	31	36.5	1315	7	US-11-063-343-35	Sequence 35, Appl
145	32	37.6	589	7	US-11-212-443-108	Sequence 108, App	218	31	36.5	1410	6	US-10-821-234-1050	Sequence 1050, Ap
146	32	37.6	589	7	US-11-212-443-109	Sequence 109, App	219	31	36.5	1467	6	US-10-507-956-1	Sequence 1, Appli
147	32	37.6	595	7	US-11-212-443-97	Sequence 97, Appl	220	31	36.5	2080	6	US-10-821-234-1640	Sequence 1640, Ap
148	32	37.6	614	7	US-11-110-748-2	Sequence 2, Appli	221	31	36.5	3433	6	US-10-714-781A-67	Sequence 67, Appl
149	32	37.6	665	6	US-10-873-528-127	Sequence 127, App	222	30.5	35.9	168	6	US-10-454-437-196	Sequence 196, App
150	32	37.6	884	7	US-11-098-686-11426	Sequence 11426, A	223	30.5	35.9	168	7	US-11-055-822-360	Sequence 360, App
151	32	37.6	1438	6	US-10-511-559-73	Sequence 73, Appl	224	30.5	35.9	225	6	US-10-209-208-1	Sequence 1, Appli
152	32	37.6	1538	7	US-11-043-889-37	Sequence 37, Appl	225	30.5	35.9	225	6	US-10-209-208-20	Sequence 20, Appl
153	32	37.6	2096	6	US-10-995-561-606	Sequence 606, App	226	30.5	35.9	225	7	US-11-218-880-1	Sequence 1, Appli
154	32	37.6	2351	6	US-10-995-561-608	Sequence 608, App	227	30.5	35.9	225	7	US-11-218-880-20	Sequence 20, Appl
155	31.5	37.1	94	6	US-10-467-657-2090	Sequence 2090, Ap	228	30.5	35.9	225	7	US-11-100-988-2	Sequence 2, Appli
156	31.5	37.1	317	7	US-11-011-332A-21	Sequence 21, Appl	229	30.5	35.9	230	7	US-11-082-154A-40	Sequence 40, Appl
157	31.5	37.1	383	7	US-11-092-168-7	Sequence 7, Appli	230	30.5	35.9	338	7	US-11-207-626A-28	Sequence 28, Appl
158	31.5	37.1	619	7	US-11-156-953-5	Sequence 5, Appli	231	30.5	35.9	508	7	US-11-024-959-467	Sequence 467, App
159	31.5	37.1	1099	7	US-11-148-770-32	Sequence 32, Appl	232	30.5	35.9	528	7	US-11-024-959-343	Sequence 343, App
160	31.5	37.1	1089	7	US-11-076-427A-14	Sequence 14, Appl	233	30.5	35.9	1037	7	US-11-052-554A-250	Sequence 250, App
161	31.5	37.1	1451	7	US-10-995-561-829	Sequence 829, App	234	30.5	35.9	1102	7	US-11-098-686-10951	Sequence 10951, A
162	31.5	37.1	2505	7	US-11-126-313-33	Sequence 33, Appl	235	30.5	35.9	2233	6	US-10-873-528-2	Sequence 2, Appli
163	31	36.5	31	6	US-10-467-657-7786	Sequence 7786, Ap	236	30	35.3	93	6	US-10-467-657-1402	Sequence 1402, Ap
164	31	36.5	93	6	US-10-467-657-8731	Sequence 8731, Ap	237	30	35.3	93	7	US-11-024-959-407	Sequence 407, App
165	31	36.5	118	6	US-10-850-635-8	Sequence 8, Appli	238	30	35.3	97	5	US-09-978-360A-695	Sequence 695, App
166	31	36.5	123	7	US-11-182-016-45	Sequence 45, Appl	239	30	35.3	105	7	US-11-024-959-405	Sequence 405, App
167	31	36.5	129	6	US-10-821-234-1250	Sequence 1250, Ap	240	30	35.3	105	7	US-11-024-959-406	Sequence 406, App
168	31	36.5	171	6	US-10-467-657-1838	Sequence 1838, Ap	241	30	35.3	129	6	US-10-501-039-8	Sequence 8, Appli
169	31	36.5	204	7	US-11-169-041-161	Sequence 161, App	242	30	35.3	149	6	US-10-467-657-8606	Sequence 8606, Ap
170	31	36.5	209	7	US-11-134-241-37	Sequence 37, Appl	243	30	35.3	158	6	US-10-510-386-128	Sequence 128, App
171	31	36.5	218	7	US-11-218-821-8	Sequence 8, Appli	244	30	35.3	160	7	US-11-044-285-1	Sequence 1, Appli

245	30	35.3	168	6	US-10-927-641-69	Sequence 69, Appl	318	30	35.3	551	6	US-10-886-504-5	Sequence 5, Appl
246	30	35.3	215	6	US-10-374-954-9	Sequence 9, Appl	319	30	35.3	551	6	US-10-886-505-5	Sequence 5, Appl
247	30	35.3	216	6	US-10-798-579A-2	Sequence 2, Appl	320	30	35.3	551	6	US-10-886-527-5	Sequence 5, Appl
248	30	35.3	216	6	US-10-883-512-95	Sequence 95, Appl	321	30	35.3	554	6	US-11-074-176-320	Sequence 320, App
249	30	35.3	216	6	US-10-883-512-96	Sequence 96, Appl	322	30	35.3	556	6	US-10-063-703-24	Sequence 24, Appl
250	30	35.3	216	6	US-10-838-616-59	Sequence 59, Appl	323	30	35.3	556	7	US-11-102-240-24	Sequence 20, Appl
251	30	35.3	216	6	US-10-714-887-426	Sequence 426, App	324	30	35.3	558	7	US-11-052-554A-20	Sequence 20, Appl
252	30	35.3	216	6	US-10-887-475B-2	Sequence 2, Appl	325	30	35.3	558	7	US-11-134-563-12	Sequence 12, Appl
253	30	35.3	216	7	US-11-114-672-12	Sequence 12, Appl	326	30	35.3	582	7	US-11-074-176-68	Sequence 68, Appl
254	30	35.3	228	6	US-10-793-626-1862	Sequence 1862, Ap	327	30	35.3	582	7	US-11-074-176-110	Sequence 110, App
255	30	35.3	230	6	US-10-884-730-373	Sequence 373, App	328	30	35.3	599	6	US-11-212-443-100	Sequence 100, App
256	30	35.3	235	7	US-11-188-473-2	Sequence 2, Appl	329	30	35.3	600	6	US-10-606-302-3	Sequence 3, Appl
257	30	35.3	238	6	US-10-927-641-68	Sequence 68, Appl	330	30	35.3	615	6	US-10-524-647-134	Sequence 134, App
258	30	35.3	265	7	US-11-037-243-102	Sequence 102, App	331	30	35.3	625	6	US-10-793-626-2464	Sequence 2464, App
259	30	35.3	289	6	US-10-878-556A-122	Sequence 122, App	332	30	35.3	668	6	US-10-995-561-619	Sequence 619, App
260	30	35.3	301	6	US-10-793-626-1826	Sequence 1826, Ap	333	30	35.3	686	7	US-11-144-630-71	Sequence 71, Appl
261	30	35.3	305	6	US-10-793-626-1362	Sequence 1362, Ap	334	30	35.3	691	6	US-10-995-561-617	Sequence 617, App
262	30	35.3	314	7	US-11-108-172-1110	Sequence 1110, Ap	335	30	35.3	697	6	US-10-485-517-202	Sequence 202, App
263	30	35.3	323	7	US-11-140-625-9	Sequence 9, Appl	336	30	35.3	699	7	US-11-130-206-4	Sequence 4, Appl
264	30	35.3	333	6	US-10-131-826A-132	Sequence 132, App	337	30	35.3	700	7	US-11-130-206-2	Sequence 2, Appl
265	30	35.3	339	6	US-10-485-517-367	Sequence 367, App	338	30	35.3	731	7	US-11-098-686-10415	Sequence 10415, A
266	30	35.3	341	6	US-10-467-657-5408	Sequence 5408, Ap	339	30	35.3	742	6	US-10-995-561-615	Sequence 615, App
267	30	35.3	348	6	US-10-667-295-121	Sequence 121, App	340	30	35.3	742	6	US-10-995-561-618	Sequence 618, App
268	30	35.3	350	6	US-10-667-295-120	Sequence 120, App	341	30	35.3	742	7	US-11-169-041-184	Sequence 184, App
269	30	35.3	354	6	US-10-821-234-1198	Sequence 1198, Ap	342	30	35.3	743	6	US-10-467-657-4082	Sequence 4082, Ap
270	30	35.3	375	7	US-10-667-295-119	Sequence 119, App	343	30	35.3	770	6	US-10-982-545-15	Sequence 15, Appl
271	30	35.3	377	7	US-11-019-711-66	Sequence 66, Appl	344	30	35.3	770	6	US-10-789-273-38	Sequence 38, Appl
272	30	35.3	384	6	US-10-513-759-16	Sequence 16, Appl	345	30	35.3	808	6	US-10-878-556A-80	Sequence 80, Appl
273	30	35.3	429	6	US-10-886-504-2	Sequence 2, Appl	346	30	35.3	809	6	US-10-467-657-1106	Sequence 1106, Ap
274	30	35.3	429	6	US-10-886-504-3	Sequence 3, Appl	347	30	35.3	810	7	US-11-052-554A-245	Sequence 245, App
275	30	35.3	429	6	US-10-886-504-4	Sequence 4, Appl	348	30	35.3	813	7	US-11-098-686-10839	Sequence 10839, A
276	30	35.3	429	6	US-10-886-504-13	Sequence 13, Appl	349	30	35.3	818	6	US-10-501-675-9	Sequence 9, Appl
277	30	35.3	429	6	US-10-886-505-2	Sequence 2, Appl	350	30	35.3	824	6	US-10-957-569-31	Sequence 31, Appl
278	30	35.3	429	6	US-10-886-505-3	Sequence 3, Appl	351	30	35.3	824	7	US-11-097-589-29	Sequence 29, Appl
279	30	35.3	429	6	US-10-886-505-4	Sequence 4, Appl	352	30	35.3	829	6	US-10-512-109-27	Sequence 27, Appl
280	30	35.3	429	6	US-10-886-505-13	Sequence 13, Appl	353	30	35.3	858	6	US-10-613-744-6	Sequence 6, Appl
281	30	35.3	429	6	US-10-886-527-2	Sequence 2, Appl	354	30	35.3	873	7	US-11-057-058-58	Sequence 58, Appl
282	30	35.3	429	6	US-10-886-527-3	Sequence 3, Appl	355	30	35.3	919	6	US-10-858-730-206	Sequence 206, App
283	30	35.3	429	6	US-10-886-527-4	Sequence 4, Appl	356	30	35.3	979	6	US-10-636-320-6	Sequence 6, Appl
284	30	35.3	429	6	US-10-886-527-13	Sequence 13, Appl	357	30	35.3	1029	6	US-10-821-234-908	Sequence 908, App
285	30	35.3	432	7	US-11-055-822-224	Sequence 224, App	358	30	35.3	1065	6	US-10-793-626-1212	Sequence 1212, App
286	30	35.3	432	7	US-11-000-463-365	Sequence 365, App	359	30	35.3	1160	6	US-10-131-826A-234	Sequence 234, App
287	30	35.3	433	6	US-10-131-826A-6	Sequence 6, Appl	360	30	35.3	1162	6	US-10-451-375-3	Sequence 3, Appl
288	30	35.3	453	7	US-11-098-686-10358	Sequence 10358, A	361	30	35.3	1198	6	US-10-451-375-4	Sequence 4, Appl
289	30	35.3	458	7	US-11-000-463-350	Sequence 350, App	362	30	35.3	1439	7	US-11-124-368A-291	Sequence 291, App
290	30	35.3	458	7	US-11-000-463-822	Sequence 822, App	363	30	35.3	1454	7	US-11-109-157A-2	Sequence 2, Appl
291	30	35.3	478	7	US-11-055-822-220	Sequence 220, App	364	30	35.3	1686	7	US-11-109-157A-1	Sequence 1, Appl
292	30	35.3	492	6	US-10-467-657-1804	Sequence 1804, Ap	365	30	35.3	1686	7	US-11-226-701-2	Sequence 2, Appl
293	30	35.3	527	6	US-10-886-504-7	Sequence 7, Appl	366	29.5	34.7	15	7	US-11-045-024-12973	Sequence 12973, A
294	30	35.3	527	6	US-10-886-504-8	Sequence 8, Appl	367	29.5	34.7	143	6	US-10-793-626-1714	Sequence 1714, Ap
295	30	35.3	527	6	US-10-886-504-9	Sequence 9, Appl	368	29.5	34.7	199	6	US-10-401-386B-29	Sequence 29, Appl
296	30	35.3	527	6	US-10-886-504-10	Sequence 10, Appl	369	29.5	34.7	199	6	US-10-511-988-14	Sequence 14, Appl
297	30	35.3	527	6	US-10-886-505-7	Sequence 7, Appl	370	29.5	34.7	206	6	US-10-511-988-12	Sequence 12, Appl
298	30	35.3	527	6	US-10-886-505-8	Sequence 8, Appl	371	29.5	34.7	223	6	US-10-511-988-10	Sequence 10, Appl
299	30	35.3	527	6	US-10-886-505-9	Sequence 9, Appl	372	29.5	34.7	230	7	US-11-098-686-11128	Sequence 11128, A
300	30	35.3	527	6	US-10-886-505-10	Sequence 10, Appl	373	29.5	34.7	246	6	US-10-485-517-242	Sequence 242, App
301	30	35.3	527	6	US-10-886-527-7	Sequence 7, Appl	374	29.5	34.7	261	6	US-10-055-877-164	Sequence 164, App
302	30	35.3	527	6	US-10-886-527-8	Sequence 8, Appl	375	29.5	34.7	261	6	US-10-055-877-165	Sequence 165, App
303	30	35.3	527	6	US-10-886-527-9	Sequence 9, Appl	376	29.5	34.7	350	6	US-10-485-517-209	Sequence 209, App
304	30	35.3	527	6	US-10-886-527-10	Sequence 10, Appl	377	29.5	34.7	486	7	US-11-108-172-685	Sequence 685, App
305	30	35.3	530	6	US-10-886-504-1	Sequence 1, Appl	378	29.5	34.7	532	6	US-10-467-657-820	Sequence 820, App
306	30	35.3	530	6	US-10-886-505-1	Sequence 1, Appl	379	29.5	34.7	1141	6	US-10-601-368-6	Sequence 6, Appl
307	30	35.3	530	6	US-10-886-527-11	Sequence 11, Appl	380	29.5	34.7	1141	6	US-10-601-368-24	Sequence 24, Appl
308	30	35.3	536	7	US-11-156-003-14	Sequence 14, Appl	381	29.5	34.7	1166	6	US-10-601-368-4	Sequence 4, Appl
309	30	35.3	536	7	US-11-156-003-15	Sequence 15, Appl	382	29.5	34.7	1166	6	US-10-601-368-22	Sequence 22, Appl
310	30	35.3	542	7	US-11-156-003-1	Sequence 1, Appl	383	29.5	34.7	1188	6	US-10-601-368-3	Sequence 3, Appl
311	30	35.3	542	7	US-11-156-003-10	Sequence 10, Appl	384	29.5	34.7	1188	6	US-10-601-368-21	Sequence 21, Appl
312	30	35.3	542	7	US-11-156-003-11	Sequence 11, Appl	385	29.5	34.7	1188	7	US-11-000-463-338	Sequence 338, App
313	30	35.3	542	7	US-11-156-003-12	Sequence 12, Appl	386	29.5	34.7	1188	7	US-11-000-463-810	Sequence 810, App
314	30	35.3	542	7	US-11-156-003-13	Sequence 13, Appl	387	29.5	34.7	2036	7	US-11-124-368A-276	Sequence 276, App
315	30	35.3	548	6	US-10-886-504-11	Sequence 11, Appl	388	29.5	34.7	2036	7	US-11-124-368A-280	Sequence 280, App
316	30	35.3	548	6	US-10-886-505-11	Sequence 11, Appl	389	29.5	34.7	2036	7	US-11-124-368A-281	Sequence 281, App
317	30	35.3	548	6	US-10-886-527-11	Sequence 11, Appl	390	29	34.1	16	7	US-11-167-636-33	Sequence 33, Appl

391	29	34.1	19	6	US-10-503-575-118	Sequence 118, App	464	29	34.1	465	7	US-11-098-686-10877	Sequence 10877, A
392	29	34.1	42	6	US-10-467-657-4636	Sequence 4636, Ap	465	29	34.1	472	6	US-10-485-517-153	Sequence 153, App
393	29	34.1	56	6	US-10-467-657-3070	Sequence 3070, Ap	466	29	34.1	476	6	US-11-024-959-519	Sequence 519, App
394	29	34.1	69	6	US-11-237-597-42	Sequence 42, Appl	467	29	34.1	477	6	US-10-877-346-47	Sequence 47, Appl
395	29	34.1	81	6	US-10-511-122-18	Sequence 18, Appl	468	29	34.1	487	6	US-10-467-657-2272	Sequence 2272, Ap
396	29	34.1	81	6	US-10-511-122-18	Sequence 18, Appl	469	29	34.1	491	6	US-11-069-642-137	Sequence 137, App
397	29	34.1	101	6	US-10-816-768-49	Sequence 49, Appl	470	29	34.1	497	7	US-11-149-462-11	Sequence 11, Appl
398	29	34.1	101	6	US-11-226-555-9	Sequence 9, Appl	471	29	34.1	497	7	US-11-049-536-163	Sequence 163, App
399	29	34.1	132	7	US-11-088-686-35	Sequence 35, Appl	472	29	34.1	510	6	US-10-467-657-3260	Sequence 3260, Ap
400	29	34.1	141	7	US-11-098-686-10315	Sequence 10315, A	473	29	34.1	512	6	US-10-517-939-236	Sequence 236, App
401	29	34.1	142	6	US-10-995-951A-26	Sequence 26, Appl	474	29	34.1	514	7	US-11-195-739-1	Sequence 1, Appl
402	29	34.1	142	6	US-10-995-951A-29	Sequence 29, Appl	475	29	34.1	529	7	US-11-132-947-8	Sequence 8, Appl
403	29	34.1	142	7	US-11-067-425A-1	Sequence 1, Appl	476	29	34.1	534	6	US-10-793-626-920	Sequence 920, App
404	29	34.1	142	7	US-11-067-425A-64	Sequence 64, Appl	477	29	34.1	543	7	US-11-113-224-6	Sequence 6, Appl
405	29	34.1	143	6	US-10-821-234-1077	Sequence 1077, Ap	478	29	34.1	543	7	US-11-130-391-3	Sequence 3, Appl
406	29	34.1	149	6	US-10-995-951A-27	Sequence 27, Appl	479	29	34.1	554	7	US-11-098-686-10420	Sequence 10420, A
407	29	34.1	149	7	US-11-067-425A-4	Sequence 4, Appl	480	29	34.1	563	6	US-10-454-437-120	Sequence 120, App
408	29	34.1	152	7	US-11-059-814-16	Sequence 16, Appl	481	29	34.1	576	6	US-10-467-657-8146	Sequence 8146, Ap
409	29	34.1	155	6	US-10-467-657-7764	Sequence 7764, Ap	482	29	34.1	579	7	US-11-045-802-32	Sequence 32, Appl
410	29	34.1	157	6	US-10-467-657-5496	Sequence 5496, Ap	483	29	34.1	579	7	US-11-045-802-33	Sequence 33, Appl
411	29	34.1	167	6	US-10-793-626-132	Sequence 132, App	484	29	34.1	591	6	US-10-453-372-952	Sequence 952, App
412	29	34.1	181	6	US-10-055-877-75	Sequence 75, Appl	485	29	34.1	592	7	US-11-059-292A-10	Sequence 10, Appl
413	29	34.1	205	6	US-10-821-234-968	Sequence 968, App	486	29	34.1	605	6	US-10-467-657-3946	Sequence 3946, Ap
414	29	34.1	208	7	US-11-088-686-39	Sequence 39, Appl	487	29	34.1	654	6	US-10-510-947-6	Sequence 6, Appl
415	29	34.1	216	6	US-10-467-657-1284	Sequence 1284, Ap	488	29	34.1	662	6	US-10-493-909-80	Sequence 80, Appl
416	29	34.1	216	6	US-10-467-657-5532	Sequence 5532, Ap	489	29	34.1	667	7	US-10-821-234-1477	Sequence 1477, Ap
417	29	34.1	216	6	US-10-467-657-7492	Sequence 7492, Ap	490	29	34.1	667	7	US-11-124-368A-239	Sequence 239, App
418	29	34.1	216	6	US-10-467-657-8232	Sequence 8232, Ap	491	29	34.1	668	6	US-11-124-368A-250	Sequence 250, App
419	29	34.1	216	6	US-10-467-657-8310	Sequence 8310, Ap	492	29	34.1	682	7	US-11-039-756-2	Sequence 2, Appl
420	29	34.1	218	7	US-11-218-821-9	Sequence 9, Appl	493	29	34.1	683	7	US-11-124-368A-243	Sequence 243, App
421	29	34.1	225	7	US-11-055-822-990	Sequence 990, App	494	29	34.1	686	7	US-11-124-368A-239	Sequence 239, App
422	29	34.1	237	7	US-11-108-172-1062	Sequence 1062, Ap	495	29	34.1	688	7	US-11-113-424-45	Sequence 45, Appl
423	29	34.1	243	7	US-11-108-172-1122	Sequence 1122, Ap	496	29	34.1	688	7	US-11-113-424-48	Sequence 48, Appl
424	29	34.1	268	6	US-10-467-657-4064	Sequence 4064, Ap	497	29	34.1	688	7	US-11-113-424-48	Sequence 48, Appl
425	29	34.1	274	6	US-10-454-437-354	Sequence 354, App	498	29	34.1	726	7	US-11-113-424-46	Sequence 46, Appl
426	29	34.1	280	5	US-09-978-360A-809	Sequence 809, App	499	29	34.1	727	6	US-11-124-368A-247	Sequence 247, App
427	29	34.1	282	7	US-11-058-924-9	Sequence 9, Appl	500	29	34.1	727	6	US-10-995-561-864	Sequence 864, App
428	29	34.1	294	7	US-11-043-788-165	Sequence 165, App	501	29	34.1	728	7	US-11-124-635-6	Sequence 6, Appl
429	29	34.1	294	7	US-11-052-554A-71	Sequence 71, Appl	502	29	34.1	728	7	US-11-124-635-8	Sequence 8, Appl
430	29	34.1	296	7	US-11-156-084-248	Sequence 248, App	503	29	34.1	728	7	US-11-043-788-164	Sequence 164, App
431	29	34.1	296	7	US-11-156-084-285	Sequence 285, App	504	29	34.1	732	7	US-11-124-368A-248	Sequence 248, App
432	29	34.1	296	7	US-11-043-788-168	Sequence 168, App	505	29	34.1	740	7	US-11-137-465-61	Sequence 61, Appl
433	29	34.1	308	6	US-10-131-826A-100	Sequence 100, Appl	506	29	34.1	750	7	US-11-073-347-1	Sequence 1, Appl
434	29	34.1	313	6	US-10-723-207-3	Sequence 3, Appl	507	29	34.1	750	7	US-11-124-368A-244	Sequence 244, App
435	29	34.1	317	6	US-10-995-561-828	Sequence 828, App	508	29	34.1	750	7	US-11-124-368A-249	Sequence 249, App
436	29	34.1	317	6	US-11-186-284-6	Sequence 6, Appl	509	29	34.1	750	7	US-11-155-288-4	Sequence 4, Appl
437	29	34.1	317	7	US-11-055-309A-4	Sequence 4, Appl	510	29	34.1	750	7	US-11-202-516-2	Sequence 2, Appl
438	29	34.1	317	7	US-11-055-309A-5	Sequence 5, Appl	511	29	34.1	753	7	US-11-124-368A-237	Sequence 237, App
439	29	34.1	317	7	US-11-055-309A-6	Sequence 6, Appl	512	29	34.1	773	7	US-11-169-041-213	Sequence 213, App
440	29	34.1	324	6	US-10-511-314-19	Sequence 19, Appl	513	29	34.1	777	7	US-11-124-368A-238	Sequence 238, App
441	29	34.1	324	6	US-10-511-314-19	Sequence 19, Appl	514	29	34.1	778	7	US-11-124-368A-240	Sequence 240, App
442	29	34.1	327	7	US-11-156-084-238	Sequence 238, App	515	29	34.1	782	6	US-10-995-561-861	Sequence 861, App
443	29	34.1	333	7	US-11-008-570-39	Sequence 39, Appl	516	29	34.1	791	7	US-11-124-368A-245	Sequence 245, App
444	29	34.1	337	6	US-10-467-657-4674	Sequence 4674, Ap	517	29	34.1	792	7	US-11-088-686-29	Sequence 29, Appl
445	29	34.1	344	6	US-11-108-172-1085	Sequence 1085, Ap	518	29	34.1	800	6	US-10-510-716-7	Sequence 7, Appl
446	29	34.1	365	6	US-10-821-234-1575	Sequence 1575, Ap	519	29	34.1	802	6	US-10-510-386-2	Sequence 2, Appl
447	29	34.1	377	7	US-11-124-103-2	Sequence 2, Appl	520	29	34.1	803	7	US-11-124-368A-241	Sequence 241, App
448	29	34.1	380	7	US-11-176-868-13	Sequence 13, Appl	521	29	34.1	803	7	US-11-124-368A-242	Sequence 242, App
449	29	34.1	389	7	US-11-088-686-1	Sequence 1, Appl	522	29	34.1	810	6	US-10-510-716-9	Sequence 9, Appl
450	29	34.1	389	7	US-11-088-686-3	Sequence 3, Appl	523	29	34.1	811	7	US-11-055-822-1142	Sequence 1142, Ap
451	29	34.1	389	7	US-11-088-686-5	Sequence 5, Appl	524	29	34.1	811	7	US-11-055-822-1144	Sequence 1144, Ap
452	29	34.1	389	7	US-11-088-686-7	Sequence 7, Appl	525	29	34.1	811	7	US-11-124-291-4	Sequence 4, Appl
453	29	34.1	389	7	US-11-088-686-9	Sequence 9, Appl	526	29	34.1	817	7	US-11-144-630-68	Sequence 68, Appl
454	29	34.1	389	7	US-11-088-686-11	Sequence 11, Appl	527	29	34.1	822	6	US-10-510-716-8	Sequence 8, Appl
455	29	34.1	389	7	US-11-088-686-13	Sequence 13, Appl	528	29	34.1	824	6	US-10-510-716-2	Sequence 2, Appl
456	29	34.1	390	7	US-11-194-246-288	Sequence 288, App	529	29	34.1	836	7	US-11-124-368A-246	Sequence 246, App
457	29	34.1	395	5	US-09-978-360A-769	Sequence 769, App	530	29	34.1	847	6	US-10-995-561-863	Sequence 863, App
458	29	34.1	396	7	US-11-051-568-5	Sequence 5, Appl	531	29	34.1	847	6	US-10-995-561-865	Sequence 865, App
459	29	34.1	406	7	US-11-000-463-377	Sequence 377, App	532	29	34.1	864	7	US-11-053-100-58	Sequence 58, Appl
460	29	34.1	436	6	US-11-000-463-849	Sequence 404, App	533	29	34.1	879	7	US-11-124-367A-329	Sequence 329, App
461	29	34.1	436	6	US-10-131-826A-404	Sequence 404, App	534	29	34.1	910	6	US-10-131-826A-112	Sequence 112, App
462	29	34.1	443	7	US-11-112-882-29	Sequence 29, Appl	535	29	34.1	912	7	US-11-124-367A-328	Sequence 328, App
463	29	34.1	446	7	US-11-108-172-1121	Sequence 1121, Ap	536	29	34.1	992	7	US-11-098-686-10761	Sequence 10761, A



537	29	34.1	1007	6	US-10-467-657-8514	Sequence 8514, App	610	28	32.9	217	6	US-10-467-657-4802	Sequence 4802, App
538	29	34.1	1049	7	US-11-137-465-42	Sequence 42, Appl	611	28	32.9	217	6	US-10-467-657-6518	Sequence 6518, App
539	29	34.1	1170	7	US-11-080-026-2	Sequence 2, Appl	612	28	32.9	217	6	US-10-467-657-8044	Sequence 8044, App
540	29	34.1	1170	7	US-11-107-028-4	Sequence 4, Appl	613	28	32.9	218	6	US-10-983-120-19	Sequence 19, Appl
541	29	34.1	1250	7	US-11-137-465-62	Sequence 62, Appl	614	28	32.9	219	6	US-10-467-657-4956	Sequence 4956, App
542	29	34.1	1288	7	US-11-052-554A-93	Sequence 93, Appl	615	28	32.9	219	7	US-11-145-861-9	Sequence 9, Appl
543	29	34.1	1296	6	US-10-615-668-3	Sequence 3, Appl	616	28	32.9	222	7	US-11-125-263A-8	Sequence 8, Appl
544	29	34.1	1307	6	US-10-995-561-711	Sequence 711, App	617	28	32.9	225	6	US-10-209-208-79	Sequence 79, Appl
545	29	34.1	1402	6	US-10-971-982-2	Sequence 2, Appl	618	28	32.9	225	7	US-11-053-076-272	Sequence 372, App
546	29	34.1	1538	6	US-10-995-561-772	Sequence 772, App	619	28	32.9	229	7	US-11-063-343-34	Sequence 34, Appl
547	29	34.1	1591	6	US-10-495-083-5	Sequence 5, Appl	620	28	32.9	232	7	US-11-052-554A-186	Sequence 186, App
548	29	34.1	1607	7	US-11-098-686-10178	Sequence 10178, A	621	28	32.9	232	7	US-11-194-890-12	Sequence 12, Appl
549	29	34.1	1857	7	US-11-057-058-60	Sequence 60, Appl	622	28	32.9	247	7	US-11-124-368A-275	Sequence 275, App
550	29	34.1	1857	7	US-11-057-058-61	Sequence 61, Appl	623	28	32.9	255	7	US-11-054-515-1626	Sequence 1626, App
551	29	34.1	2011	7	US-11-080-991-56	Sequence 56, Appl	624	28	32.9	257	7	US-11-096-051-12	Sequence 12, Appl
552	29	34.1	2725	7	US-11-113-424-52	Sequence 52, Appl	625	28	32.9	258	6	US-10-821-243-1024	Sequence 1024, App
553	29	34.1	3655	7	US-11-075-185-5	Sequence 5, Appl	626	28	32.9	261	7	US-11-037-243-78	Sequence 78, Appl
554	29	34.1	5335	6	US-10-995-561-777	Sequence 777, App	627	28	32.9	264	7	US-11-165-067A-28	Sequence 28, Appl
555	29	34.1	5405	6	US-10-995-561-774	Sequence 774, App	628	28	32.9	264	7	US-11-165-067A-45	Sequence 45, Appl
556	29	34.1	5415	6	US-10-995-561-779	Sequence 779, App	629	28	32.9	267	7	US-11-194-890-11	Sequence 11, Appl
557	29	34.1	5464	6	US-10-995-561-775	Sequence 775, App	630	28	32.9	272	6	US-10-793-626-520	Sequence 520, App
558	29	34.1	5935	6	US-10-995-561-776	Sequence 776, App	631	28	32.9	279	6	US-10-467-657-8266	Sequence 8266, App
559	28.5	33.5	138	6	US-10-467-657-8342	Sequence 8342, App	632	28	32.9	288	7	US-11-029-188-2	Sequence 2, Appl
560	28.5	33.5	153	6	US-10-467-657-4546	Sequence 4546, App	633	28	32.9	291	6	US-10-995-561-558	Sequence 558, App
561	28.5	33.5	250	7	US-11-052-554A-293	Sequence 293, App	634	28	32.9	292	6	US-10-527-500-9	Sequence 9, Appl
562	28.5	33.5	250	7	US-11-052-554A-294	Sequence 294, App	635	28	32.9	293	6	US-10-131-826A-422	Sequence 422, App
563	28.5	33.5	278	6	US-10-793-626-2208	Sequence 2208, App	636	28	32.9	293	6	US-10-467-657-4834	Sequence 4834, App
564	28.5	33.5	284	6	US-10-131-826A-372	Sequence 372, App	637	28	32.9	293	6	US-10-995-561-562	Sequence 562, App
565	28.5	33.5	284	7	US-11-170-268-40	Sequence 40, Appl	638	28	32.9	295	6	US-10-485-517-256	Sequence 256, App
566	28.5	33.5	341	7	US-11-170-268-28	Sequence 28, Appl	639	28	32.9	296	7	US-11-096-051-18	Sequence 18, Appl
567	28.5	33.5	341	7	US-11-170-268-36	Sequence 36, Appl	640	28	32.9	299	7	US-11-074-176-198	Sequence 198, App
568	28.5	33.5	350	7	US-11-024-959-379	Sequence 379, App	641	28	32.9	313	7	US-11-098-686-10326	Sequence 10326, A
569	28.5	33.5	378	6	US-11-207-626A-34	Sequence 34, Appl	642	28	32.9	313	7	US-11-237-600-38	Sequence 68, Appl
570	28.5	33.5	403	6	US-10-873-528-32	Sequence 32, Appl	643	28	32.9	316	6	US-10-995-793-77	Sequence 77, Appl
571	28.5	33.5	556	7	US-11-154-227-98	Sequence 98, Appl	644	28	32.9	316	7	US-11-125-263A-4	Sequence 4, Appl
572	28.5	33.5	574	7	US-11-120-308-166	Sequence 166, App	645	28	32.9	316	7	US-11-156-084-193	Sequence 193, App
573	28.5	33.5	585	6	US-10-467-657-3160	Sequence 3160, App	646	28	32.9	316	7	US-11-156-084-210	Sequence 210, App
574	28.5	33.5	601	6	US-11-024-959-295	Sequence 295, App	647	28	32.9	316	7	US-11-156-084-280	Sequence 280, App
575	28.5	33.5	672	6	US-10-689-742-70	Sequence 70, Appl	648	28	32.9	316	7	US-11-156-084-287	Sequence 287, App
576	28.5	33.5	862	7	US-11-182-016-35	Sequence 35, Appl	649	28	32.9	317	6	US-10-828-033-30	Sequence 30, Appl
577	28.5	33.5	944	7	US-11-057-058-68	Sequence 68, Appl	650	28	32.9	317	7	US-11-124-368A-272	Sequence 272, App
578	28.5	33.5	966	7	US-11-057-058-67	Sequence 67, Appl	651	28	32.9	317	7	US-11-124-368A-274	Sequence 274, App
579	28.5	33.5	1637	6	US-10-821-234-1204	Sequence 1204, App	652	28	32.9	319	6	US-10-454-437-114	Sequence 114, App
580	28.5	33.5	1706	7	US-11-052-554A-176	Sequence 176, App	653	28	32.9	319	6	US-10-454-437-116	Sequence 116, App
581	28.5	33.5	2105	7	US-11-052-554A-173	Sequence 173, App	654	28	32.9	321	6	US-10-063-703-10	Sequence 10, Appl
582	28	32.9	14	6	US-10-723-207-46	Sequence 46, Appl	655	28	32.9	321	7	US-11-102-240-10	Sequence 10, Appl
583	28	32.9	25	6	US-10-939-890-300	Sequence 300, App	656	28	32.9	327	6	US-10-821-234-884	Sequence 884, App
584	28	32.9	53	7	US-11-055-822-530	Sequence 530, App	657	28	32.9	330	6	US-10-467-657-2798	Sequence 2798, App
585	28	32.9	64	6	US-10-467-657-3224	Sequence 3224, App	658	28	32.9	331	6	US-10-131-826A-184	Sequence 184, App
586	28	32.9	66	6	US-10-995-561-801	Sequence 801, App	659	28	32.9	336	7	US-11-205-109-4	Sequence 4, Appl
587	28	32.9	67	6	US-10-467-657-6428	Sequence 6428, App	660	28	32.9	346	7	US-11-084-624-16	Sequence 16, Appl
588	28	32.9	68	6	US-10-821-234-855	Sequence 855, App	661	28	32.9	346	7	US-11-123-013-8	Sequence 8, Appl
589	28	32.9	93	6	US-10-467-657-8861	Sequence 8861, App	662	28	32.9	346	7	US-11-169-041-208	Sequence 208, App
590	28	32.9	94	6	US-10-467-657-3902	Sequence 3902, App	663	28	32.9	359	6	US-10-201-525-11	Sequence 11, Appl
591	28	32.9	99	6	US-10-467-657-2852	Sequence 2852, App	664	28	32.9	359	7	US-11-124-368A-273	Sequence 273, App
592	28	32.9	102	7	US-11-118-855-15	Sequence 15, Appl	665	28	32.9	361	6	US-10-485-517-125	Sequence 125, App
593	28	32.9	104	6	US-10-689-742-206	Sequence 206, App	666	28	32.9	361	6	US-10-485-517-295	Sequence 295, App
594	28	32.9	136	7	US-11-055-822-966	Sequence 966, App	667	28	32.9	361	7	US-11-116-144-29	Sequence 29, Appl
595	28	32.9	136	7	US-11-055-822-994	Sequence 994, App	668	28	32.9	368	7	US-11-000-463-916	Sequence 916, App
596	28	32.9	140	5	US-09-978-360A-683	Sequence 683, App	669	28	32.9	377	6	US-10-821-234-1525	Sequence 1525, App
597	28	32.9	146	7	US-11-176-830-295	Sequence 295, App	670	28	32.9	379	6	US-10-331-826A-216	Sequence 216, App
598	28	32.9	154	6	US-10-821-234-1469	Sequence 1469, App	671	28	32.9	379	7	US-11-054-281-45	Sequence 45, Appl
599	28	32.9	179	7	US-11-098-686-11216	Sequence 11216, A	672	28	32.9	381	7	US-11-054-281-2	Sequence 2, Appl
600	28	32.9	180	6	US-10-467-657-962	Sequence 962, App	673	28	32.9	381	7	US-11-054-281-41	Sequence 41, Appl
601	28	32.9	200	7	US-11-127-632-4	Sequence 4, Appl	674	28	32.9	381	7	US-11-054-281-42	Sequence 42, Appl
602	28	32.9	201	6	US-10-821-234-1609	Sequence 1609, App	675	28	32.9	381	7	US-11-054-281-43	Sequence 43, Appl
603	28	32.9	202	6	US-10-793-626-1108	Sequence 1108, App	676	28	32.9	381	7	US-11-054-281-44	Sequence 44, Appl
604	28	32.9	215	6	US-10-467-657-6420	Sequence 6420, App	677	28	32.9	381	7	US-11-169-041-152	Sequence 152, App
605	28	32.9	216	7	US-11-156-084-244	Sequence 244, App	678	28	32.9	384	6	US-10-525-674-44	Sequence 44, Appl
606	28	32.9	216	7	US-11-156-084-322	Sequence 322, App	679	28	32.9	387	6	US-10-714-887-278	Sequence 278, App
607	28	32.9	217	6	US-10-467-657-986	Sequence 986, App	680	28	32.9	388	7	US-11-000-463-444	Sequence 444, App
608	28	32.9	217	6	US-10-467-657-2380	Sequence 2380, App	681	28	32.9	384	6	US-10-714-887-20	Sequence 20, Appl
609	28	32.9	217	6	US-10-467-657-4240	Sequence 4240, App	682	28	32.9	397	6	US-10-467-657-4982	Sequence 4982, App

683	28	32.9	397	7	US-11-219-282-11	Sequence 11, Appl	756	28	32.9	638	7	US-11-054-281-112	Sequence 112, App
684	28	32.9	402	7	US-11-098-686-10519	Sequence 10519, A	757	28	32.9	638	7	US-11-054-281-114	Sequence 114, App
685	28	32.9	411	6	US-10-467-657-8252	Sequence 8252, App	758	28	32.9	638	7	US-11-054-281-115	Sequence 115, App
686	28	32.9	413	6	US-10-517-939-110	Sequence 110, App	759	28	32.9	644	6	US-10-821-234-1107	Sequence 1107, App
687	28	32.9	421	6	US-10-458-870-2	Sequence 2, Appl	760	28	32.9	647	7	US-11-144-630-14	Sequence 14, Appl
688	28	32.9	422	7	US-11-110-851-4	Sequence 4, Appl	761	28	32.9	649	7	US-11-150-845-18	Sequence 18, Appl
689	28	32.9	424	7	US-11-098-686-10348	Sequence 10348, A	762	28	32.9	649	7	US-11-150-845-22	Sequence 22, Appl
690	28	32.9	428	7	US-11-127-622-2	Sequence 2, Appl	763	28	32.9	649	7	US-11-150-487-18	Sequence 18, Appl
691	28	32.9	428	7	US-11-127-877-73	Sequence 73, Appl	764	28	32.9	649	7	US-11-150-487-22	Sequence 22, Appl
692	28	32.9	435	7	US-11-150-845-26	Sequence 26, Appl	765	28	32.9	708	6	US-10-623-155-369	Sequence 369, App
693	28	32.9	435	7	US-11-150-487-26	Sequence 26, Appl	766	28	32.9	713	6	US-10-453-372-48	Sequence 48, Appl
694	28	32.9	436	7	US-11-150-845-32	Sequence 32, Appl	767	28	32.9	718	7	US-11-111-664-4	Sequence 4, Appl
695	28	32.9	436	7	US-11-150-487-32	Sequence 32, Appl	768	28	32.9	724	6	US-10-793-626-968	Sequence 968, App
696	28	32.9	437	7	US-11-150-845-28	Sequence 28, Appl	769	28	32.9	749	7	US-11-111-664-3	Sequence 3, Appl
697	28	32.9	437	7	US-11-150-845-30	Sequence 30, Appl	770	28	32.9	751	7	US-11-012-762-26	Sequence 26, Appl
698	28	32.9	437	7	US-11-150-487-28	Sequence 28, Appl	771	28	32.9	759	7	US-11-111-664-2	Sequence 2, Appl
699	28	32.9	437	7	US-11-150-487-30	Sequence 30, Appl	772	28	32.9	776	6	US-10-453-372-44	Sequence 44, Appl
700	28	32.9	439	6	US-10-467-657-3034	Sequence 3034, App	773	28	32.9	776	6	US-10-453-372-46	Sequence 46, Appl
701	28	32.9	446	6	US-10-527-500-1	Sequence 1, Appl	774	28	32.9	791	6	US-10-467-657-5014	Sequence 5014, App
702	28	32.9	448	6	US-10-763-712A-69	Sequence 69, Appl	775	28	32.9	806	7	US-11-108-172-1117	Sequence 1117, App
703	28	32.9	448	6	US-10-763-712A-112	Sequence 112, App	776	28	32.9	806	7	US-11-108-172-1118	Sequence 1118, App
704	28	32.9	448	7	US-11-124-367A-282	Sequence 282, App	777	28	32.9	806	7	US-11-079-900-1	Sequence 1, Appl
705	28	32.9	451	6	US-10-467-657-6194	Sequence 6194, App	778	28	32.9	816	7	US-11-057-058-10	Sequence 10, Appl
706	28	32.9	452	7	US-11-024-959-382	Sequence 382, App	779	28	32.9	816	7	US-11-144-630-67	Sequence 67, Appl
707	28	32.9	462	6	US-10-467-657-4348	Sequence 4348, App	780	28	32.9	821	7	US-11-096-051-6	Sequence 6, Appl
708	28	32.9	465	7	US-11-098-686-11183	Sequence 11183, A	781	28	32.9	823	6	US-10-453-372-122	Sequence 122, App
709	28	32.9	470	7	US-11-024-959-297	Sequence 297, App	782	28	32.9	823	6	US-10-453-372-124	Sequence 124, App
710	28	32.9	471	6	US-10-467-962B-10	Sequence 10, Appl	783	28	32.9	826	6	US-10-878-558A-68	Sequence 68, Appl
711	28	32.9	472	7	US-11-169-041-156	Sequence 156, App	784	28	32.9	826	6	US-10-453-372-118	Sequence 118, App
712	28	32.9	473	6	US-10-509-464-5	Sequence 5, Appl	785	28	32.9	826	6	US-10-453-372-156	Sequence 156, App
713	28	32.9	473	6	US-10-509-464-6	Sequence 6, Appl	786	28	32.9	826	6	US-10-453-372-166	Sequence 166, App
714	28	32.9	481	6	US-10-467-657-420	Sequence 420, App	787	28	32.9	826	6	US-10-453-372-172	Sequence 172, App
715	28	32.9	482	6	US-10-821-234-1413	Sequence 1413, App	788	28	32.9	826	6	US-10-453-372-174	Sequence 174, App
716	28	32.9	482	6	US-10-995-561-560	Sequence 560, App	789	28	32.9	826	6	US-10-453-372-176	Sequence 176, App
717	28	32.9	489	6	US-10-525-674-34	Sequence 34, Appl	790	28	32.9	832	6	US-10-453-372-120	Sequence 120, App
718	28	32.9	491	7	US-11-037-829A-9	Sequence 9, Appl	791	28	32.9	852	6	US-10-467-657-5504	Sequence 5004, App
719	28	32.9	502	6	US-10-821-234-1554	Sequence 1554, App	792	28	32.9	854	6	US-10-511-657-4	Sequence 4, Appl
720	28	32.9	503	7	US-11-037-829A-2	Sequence 2, Appl	793	28	32.9	862	7	US-11-096-051-20	Sequence 20, Appl
721	28	32.9	505	6	US-10-467-657-6228	Sequence 6228, App	794	28	32.9	866	7	US-11-052-554A-29	Sequence 29, Appl
722	28	32.9	516	6	US-10-995-561-559	Sequence 559, App	795	28	32.9	869	6	US-10-453-372-50	Sequence 50, Appl
723	28	32.9	526	7	US-11-098-686-10536	Sequence 10536, A	796	28	32.9	909	7	US-11-098-686-11181	Sequence 11181, A
724	28	32.9	527	6	US-10-705-633-1	Sequence 1, Appl	797	28	32.9	924	6	US-10-467-657-4290	Sequence 4290, App
725	28	32.9	527	6	US-10-705-633-2	Sequence 2, Appl	798	28	32.9	925	6	US-10-454-437-50	Sequence 50, Appl
726	28	32.9	536	6	US-10-705-633-3	Sequence 3, Appl	799	28	32.9	937	7	US-11-017-550-66	Sequence 66, Appl
727	28	32.9	536	6	US-10-517-939-322	Sequence 322, App	800	28	32.9	950	6	US-10-981-267-24	Sequence 24, Appl
728	28	32.9	537	6	US-10-312-109-9	Sequence 9, Appl	801	28	32.9	952	6	US-10-821-234-1557	Sequence 1557, App
729	28	32.9	562	6	US-10-995-561-561	Sequence 561, App	802	28	32.9	952	7	US-11-057-058-54	Sequence 54, Appl
730	28	32.9	562	7	US-11-169-041-153	Sequence 153, App	803	28	32.9	953	7	US-11-057-058-53	Sequence 53, Appl
731	28	32.9	569	6	US-10-632-150-2	Sequence 2, Appl	804	28	32.9	953	7	US-11-057-058-56	Sequence 56, Appl
732	28	32.9	569	7	US-11-073-457-2	Sequence 2, Appl	805	28	32.9	953	7	US-11-057-058-57	Sequence 57, Appl
733	28	32.9	569	7	US-11-099-691-12	Sequence 12, Appl	806	28	32.9	997	7	US-11-057-058-33	Sequence 33, Appl
734	28	32.9	569	7	US-11-073-460-2	Sequence 2, Appl	807	28	32.9	1011	7	US-11-057-058-31	Sequence 31, Appl
735	28	32.9	574	6	US-10-467-657-2548	Sequence 2548, App	808	28	32.9	1032	6	US-10-467-657-3278	Sequence 3278, App
736	28	32.9	575	7	US-11-024-959-374	Sequence 374, App	809	28	32.9	1052	7	US-11-124-367A-323	Sequence 323, App
737	28	32.9	581	7	US-11-045-802-30	Sequence 30, Appl	810	28	32.9	1061	7	US-11-000-463-347	Sequence 347, App
738	28	32.9	594	7	US-11-045-802-31	Sequence 31, Appl	811	28	32.9	1061	7	US-11-124-367A-325	Sequence 325, App
739	28	32.9	590	7	US-11-124-368A-183	Sequence 183, App	812	28	32.9	1091	7	US-11-000-463-348	Sequence 348, App
740	28	32.9	590	7	US-11-127-877-54	Sequence 54, Appl	813	28	32.9	1107	7	US-11-057-058-41	Sequence 41, Appl
741	28	32.9	595	6	US-10-510-386-240	Sequence 240, App	814	28	32.9	1122	6	US-10-995-561-705	Sequence 705, App
742	28	32.9	598	7	US-11-113-837-18	Sequence 18, Appl	815	28	32.9	1129	6	US-10-995-561-706	Sequence 706, App
743	28	32.9	611	6	US-10-454-437-54	Sequence 54, Appl	816	28	32.9	1139	7	US-11-098-686-10206	Sequence 10206, A
744	28	32.9	612	6	US-10-518-018-1	Sequence 1, Appl	817	28	32.9	1183	7	US-11-115-639-13	Sequence 13, Appl
745	28	32.9	614	7	US-11-052-554A-44	Sequence 44, Appl	818	28	32.9	1183	7	US-11-115-639-14	Sequence 14, Appl
746	28	32.9	614	7	US-11-150-845-34	Sequence 34, Appl	819	28	32.9	1183	7	US-11-115-639-15	Sequence 15, Appl
747	28	32.9	614	7	US-11-150-487-34	Sequence 34, Appl	820	28	32.9	1183	7	US-11-115-639-16	Sequence 16, Appl
748	28	32.9	617	7	US-11-150-845-46	Sequence 46, Appl	821	28	32.9	1183	7	US-11-115-639-17	Sequence 17, Appl
749	28	32.9	617	7	US-11-150-487-40	Sequence 40, Appl	822	28	32.9	1183	7	US-11-115-639-18	Sequence 18, Appl
750	28	32.9	618	7	US-11-033-116-2	Sequence 2, Appl	823	28	32.9	1234	6	US-10-467-657-4224	Sequence 4224, App
751	28	32.9	627	6	US-10-873-528-191	Sequence 191, App	824	28	32.9	1258	7	US-11-121-438-16	Sequence 16, Appl
752	28	32.9	628	6	US-10-467-657-2484	Sequence 2484, App	825	28	32.9	1369	7	US-11-124-367A-311	Sequence 311, App
753	28	32.9	638	6	US-10-995-561-536	Sequence 536, App	826	28	32.9	1388	6	US-10-821-234-1443	Sequence 1143, App
754	28	32.9	638	7	US-11-054-281-30	Sequence 30, Appl	827	28	32.9	1419	7	US-11-124-367A-324	Sequence 324, App
755	28	32.9	638	7	US-11-054-281-111	Sequence 111, App	828	28	32.9	1428	7	US-11-124-367A-320	Sequence 320, App

829	28	32.9	1463	7	US-11-080-991-22	Sequence 22, Appl	902	27	31.8	73	6	US-10-467-657-5022	Sequence 5022, Ap
830	28	32.9	1614	7	US-11-108-928-82	Sequence 82, Appl	903	27	31.8	79	6	US-10-770-726-58	Sequence 58, Appl
831	28	32.9	1625	6	US-10-757-832-2	Sequence 2, Appl	904	27	31.8	80	6	US-10-467-657-1988	Sequence 1988, Ap
832	28	32.9	1897	6	US-10-821-234-1635	Sequence 2, Appl	905	27	31.8	82	6	US-10-467-657-5456	Sequence 5456, Ap
833	28	32.9	1907	6	US-11-000-463-250	Sequence 1635, Ap	906	27	31.8	94	6	US-10-467-657-2214	Sequence 2214, Ap
834	28	32.9	2098	6	US-10-055-877-253	Sequence 250, App	907	27	31.8	99	6	US-10-821-234-1702	Sequence 1702, Ap
835	28	32.9	2204	6	US-11-052-554A-174	Sequence 174, App	908	27	31.8	103	6	US-10-467-657-2150	Sequence 2150, Ap
836	28	32.9	2432	6	US-10-821-234-899	Sequence 174, App	909	27	31.8	103	6	US-10-467-657-4816	Sequence 4816, Ap
837	28	32.9	2612	6	US-10-453-372-38	Sequence 899, App	910	27	31.8	105	6	US-10-485-788A-711	Sequence 711, App
838	28	32.9	2669	6	US-10-453-372-36	Sequence 36, Appl	911	27	31.8	105	6	US-11-053-076-81	Sequence 81, Appl
839	28	32.9	3104	6	US-10-453-372-34	Sequence 34, Appl	912	27	31.8	111	6	US-10-793-626-116	Sequence 116, App
840	28	32.9	3104	6	US-10-453-372-62	Sequence 62, Appl	913	27	31.8	119	7	US-11-221-900-10	Sequence 10, Appl
841	28	32.9	3104	6	US-10-453-372-64	Sequence 64, Appl	914	27	31.8	119	7	US-11-221-900-12	Sequence 12, Appl
842	28	32.9	3130	6	US-10-453-372-42	Sequence 42, Appl	915	27	31.8	119	7	US-11-221-900-13	Sequence 13, Appl
843	28	32.9	3375	7	US-11-044-111-23	Sequence 23, Appl	916	27	31.8	119	7	US-11-221-900-14	Sequence 14, Appl
844	28	32.9	3483	6	US-10-453-372-40	Sequence 40, Appl	917	27	31.8	121	6	US-10-771-257-56	Sequence 56, Appl
845	28	32.9	3546	6	US-10-453-372-32	Sequence 32, Appl	918	27	31.8	121	6	US-11-127-677-54	Sequence 54, Appl
846	28	32.9	4374	7	US-11-128-572-2	Sequence 2, Appl	919	27	31.8	122	6	US-10-498-026-47	Sequence 47, Appl
847	28	32.9	5405	7	US-11-108-172-1116	Sequence 1116, Ap	920	27	31.8	133	6	US-10-510-386-206	Sequence 206, App
848	28	32.9	8695	7	US-11-205-109-15	Sequence 15, Appl	921	27	31.8	134	6	US-10-793-626-2842	Sequence 2842, Ap
849	28	32.9	14130	7	US-11-175-689-9	Sequence 9, Appl	922	27	31.8	135	6	US-10-793-626-2676	Sequence 2676, Ap
850	28	32.9	16990	7	US-11-175-689-7	Sequence 7, Appl	923	27	31.8	135	7	US-11-074-176-128	Sequence 128, App
851	27.5	32.4	66	7	US-11-051-481-45	Sequence 45, Appl	924	27	31.8	135	7	US-11-098-686-10908	Sequence 10908, A
852	27.5	32.4	71	7	US-11-000-463-356	Sequence 356, App	925	27	31.8	137	7	US-11-022-562-225	Sequence 225, App
853	27.5	32.4	71	7	US-11-000-463-828	Sequence 828, App	926	27	31.8	138	6	US-10-793-626-762	Sequence 762, App
854	27.5	32.4	86	7	US-11-051-481-44	Sequence 44, Appl	927	27	31.8	150	6	US-10-793-626-840	Sequence 840, App
855	27.5	32.4	107	6	US-10-877-346-70	Sequence 70, Appl	928	27	31.8	152	6	US-10-485-517-411	Sequence 411, App
856	27.5	32.4	111	6	US-10-895-064-1349	Sequence 1349, Ap	929	27	31.8	152	6	US-10-793-626-1226	Sequence 1226, Ap
857	27.5	32.4	176	7	US-11-093-746A-24	Sequence 24, Appl	930	27	31.8	155	6	US-10-793-626-716	Sequence 716, App
858	27.5	32.4	212	6	US-10-793-626-642	Sequence 642, App	931	27	31.8	155	6	US-10-530-253-23	Sequence 23, Appl
859	27.5	32.4	217	6	US-10-793-626-2556	Sequence 2556, Ap	932	27	31.8	157	6	US-10-467-657-7028	Sequence 7028, Ap
860	27.5	32.4	305	7	US-11-080-091-13	Sequence 13, Appl	933	27	31.8	185	7	US-11-176-830-168	Sequence 168, App
861	27.5	32.4	305	7	US-11-087-177-11	Sequence 11, Appl	934	27	31.8	165	7	US-11-196-067-168	Sequence 168, App
862	27.5	32.4	305	7	US-11-087-177-13	Sequence 13, Appl	935	27	31.8	166	6	US-10-467-657-3512	Sequence 3512, Ap
863	27.5	32.4	316	7	US-11-092-168-10	Sequence 10, Appl	936	27	31.8	166	7	US-11-176-830-1070	Sequence 1070, Ap
864	27.5	32.4	333	6	US-10-793-626-2890	Sequence 2890, Ap	937	27	31.8	166	7	US-11-196-067-266	Sequence 266, App
865	27.5	32.4	354	7	US-10-793-626-3098	Sequence 3098, Ap	938	27	31.8	169	6	US-10-821-234-1234	Sequence 1234, Ap
866	27.5	32.4	400	7	US-11-169-041-158	Sequence 158, App	939	27	31.8	189	6	US-10-867-168-1	Sequence 1, Appl
867	27.5	32.4	421	6	US-10-131-826A-302	Sequence 302, App	940	27	31.8	174	6	US-10-532-426-1	Sequence 1, Appl
868	27.5	32.4	440	6	US-10-821-234-1282	Sequence 1282, Ap	941	27	31.8	177	6	US-10-980-388-95	Sequence 95, Appl
869	27.5	32.4	467	6	US-10-793-626-2652	Sequence 2652, Ap	942	27	31.8	179	6	US-10-793-626-942	Sequence 942, App
870	27.5	32.4	592	7	US-11-104-110-9	Sequence 9, Appl	943	27	31.8	179	6	US-10-793-626-1738	Sequence 1738, Ap
871	27.5	32.4	592	7	US-11-104-111-29	Sequence 29, Appl	944	27	31.8	183	6	US-10-793-626-2314	Sequence 2314, Ap
872	27.5	32.4	587	7	US-11-117-169-6	Sequence 6, Appl	945	27	31.8	184	7	US-11-120-308-86	Sequence 86, Appl
873	27.5	32.4	727	7	US-11-117-169-8	Sequence 8, Appl	946	27	31.8	189	6	US-10-793-626-2692	Sequence 2692, Ap
874	27.5	32.4	828	6	US-10-501-039-2	Sequence 2, Appl	947	27	31.8	195	7	US-11-019-955-24	Sequence 24, Appl
875	27.5	32.4	877	7	US-11-074-176-322	Sequence 322, App	948	27	31.8	197	7	US-11-090-878-54	Sequence 54, Appl
876	27.5	32.4	883	7	US-11-074-176-88	Sequence 88, Appl	949	27	31.8	201	7	US-11-124-368A-265	Sequence 265, App
877	27.5	32.4	882	7	US-11-082-389-396	Sequence 396, App	950	27	31.8	210	6	US-10-467-657-6256	Sequence 6256, Ap
878	27.5	32.4	1094	6	US-10-517-939-144	Sequence 144, App	951	27	31.8	212	6	US-10-990-627-5	Sequence 5, Appl
879	27.5	32.4	1227	6	US-10-793-626-96	Sequence 96, Appl	952	27	31.8	216	6	US-10-990-627-5	Sequence 5, Appl
880	27.5	32.4	1234	6	US-10-517-939-306	Sequence 306, App	953	27	31.8	217	6	US-10-467-657-388	Sequence 388, App
881	27.5	32.4	1306	6	US-10-995-561-905	Sequence 905, App	954	27	31.8	217	6	US-10-467-657-552	Sequence 552, App
882	27.5	32.4	1356	6	US-10-995-561-904	Sequence 904, App	955	27	31.8	217	6	US-10-467-657-2356	Sequence 2356, Ap
883	27.5	32.4	1356	6	US-10-995-561-906	Sequence 904, App	956	27	31.8	217	6	US-10-467-657-3800	Sequence 3800, Ap
884	27	31.8	8	6	US-10-503-575-53	Sequence 53, Appl	957	27	31.8	217	6	US-10-467-657-3930	Sequence 3930, Ap
885	27	31.8	9	7	US-11-128-440-78	Sequence 78, Appl	958	27	31.8	217	6	US-10-467-657-4018	Sequence 4018, Ap
886	27	31.8	9	7	US-11-128-440-79	Sequence 79, Appl	959	27	31.8	217	6	US-10-467-657-4438	Sequence 4438, Ap
887	27	31.8	10	6	US-10-973-977-64	Sequence 64, Appl	960	27	31.8	217	6	US-10-467-657-4726	Sequence 4726, Ap
888	27	31.8	13	7	US-11-128-440-66	Sequence 66, Appl	961	27	31.8	217	6	US-11-188-281-18	Sequence 18, Appl
889	27	31.8	14	7	US-11-128-440-65	Sequence 65, Appl	962	27	31.8	223	6	US-10-467-657-2382	Sequence 2382, Ap
890	27	31.8	15	7	US-11-128-440-64	Sequence 64, Appl	963	27	31.8	230	6	US-10-884-730-372	Sequence 372, App
891	27	31.8	16	7	US-11-054-515-3022	Sequence 3022, Ap	964	27	31.8	230	6	US-10-884-730-375	Sequence 375, App
892	27	31.8	19	6	US-10-503-575-50	Sequence 50, Appl	965	27	31.8	230	6	US-10-884-730-376	Sequence 376, App
893	27	31.8	21	6	US-10-939-890-518	Sequence 518, App	966	27	31.8	230	6	US-10-515-720A-1	Sequence 1, Appl
894	27	31.8	21	6	US-10-939-890-575	Sequence 575, App	967	27	31.8	230	6	US-11-188-281-4	Sequence 4, Appl
895	27	31.8	21	6	US-10-939-890-608	Sequence 608, App	968	27	31.8	234	6	US-10-997-697-47	Sequence 47, Appl
896	27	31.8	21	6	US-10-939-890-436	Sequence 436, App	969	27	31.8	236	7	US-11-221-900-2	Sequence 2, Appl
897	27	31.8	28	7	US-11-196-670-60	Sequence 60, Appl	970	27	31.8	238	7	US-11-120-308-96	Sequence 96, Appl
898	27	31.8	30	6	US-10-973-977-23	Sequence 23, Appl	971	27	31.8	241	6	US-10-902-546-6	Sequence 6, Appl
899	27	31.8	40	6	US-10-957-351-210	Sequence 210, App	972	27	31.8	241	7	US-11-019-955-27	Sequence 27, Appl
900	27	31.8	44	6	US-10-957-887B-304	Sequence 304, App	973	27	31.8	241	7	US-11-036-256-63	Sequence 63, Appl
901	27	31.8	71	6	US-10-467-657-750	Sequence 750, App	974	27	31.8	245	7	US-11-032-797-2	Sequence 2, Appl

975 27 31.8 248 6 US-10-530-253-7 Sequence 7, Appli  
976 27 31.8 248 6 US-10-530-253-9 Sequence 9, Appli  
977 27 31.8 248 6 US-10-530-253-11 Sequence 11, Appl  
978 27 31.8 248 7 US-11-054-515-1700 Sequence 1700, Ap  
979 27 31.8 248 7 US-11-054-515-1790 Sequence 1790, Ap  
980 27 31.8 251 7 US-11-054-515-1790 Sequence 1802, Ap  
981 27 31.8 255 7 US-11-179-977-3 Sequence 3, Appli  
982 27 31.8 257 7 US-11-054-515-1514 Sequence 1514, Ap  
983 27 31.8 261 6 US-10-467-657-400 Sequence 400, App  
984 27 31.8 267 7 US-11-098-686-10146 Sequence 10146, A  
985 27 31.8 268 7 US-11-019-955-28 Sequence 28, Appl  
986 27 31.8 269 6 US-10-793-626-880 Sequence 880, App  
987 27 31.8 271 7 US-11-091-100-17 Sequence 17, Appl  
988 27 31.8 275 6 US-10-972-587-18 Sequence 18, Appl  
989 27 31.8 275 7 US-11-098-686-10953 Sequence 10953, A  
990 27 31.8 278 7 US-11-165-067A-5 Sequence 5, Appli  
991 27 31.8 282 6 US-10-467-657-1342 Sequence 1342, Ap  
992 27 31.8 285 7 US-11-094-519A-36 Sequence 36, Appl  
993 27 31.8 288 7 US-11-156-084-38 Sequence 38, Appl  
994 27 31.8 294 7 US-11-098-686-10623 Sequence 10623, A  
995 27 31.8 295 7 US-11-091-100-2 Sequence 2, Appli  
996 27 31.8 295 7 US-11-193-205-1 Sequence 1, Appli  
997 27 31.8 297 7 US-11-156-084-37 Sequence 37, Appl  
998 27 31.8 299 6 US-10-793-626-1192 Sequence 1192, Ap  
999 27 31.8 299 6 US-10-793-626-1768 Sequence 1768, Ap  
1000 27 31.8 303 6 US-10-878-556A-74 Sequence 74, Appl

## ALIGNMENTS

## RESULT 1

US-11-098-686-10166  
; Sequence 10166, Application US/11098686  
; Publication No. US20060024696A1

## GENERAL INFORMATION:

; APPLICANT: Kapur, Vivek and Gebhart, Connie J.  
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES  
; FILE REFERENCE: 09531-128001  
; CURRENT APPLICATION NUMBER: US/11/098,686  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: PCT/US03/31318  
; PRIOR FILING DATE: 2003-10-01  
; PRIOR APPLICATION NUMBER: US 60/416,395  
; PRIOR FILING DATE: 2002-10-04  
; NUMBER OF SEQ ID NOS: 11433  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10166  
; LENGTH: 179  
; TYPE: PRT  
; ORGANISM: Lawsonia intracellularis

US-11-098-686-10166

Query Match 52.9%; Score 45; DB 7; Length 179;  
Best Local Similarity 53.8%; Pred. No. 1.7;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 1 QWGESEKPYDDH 13  
|||||:|:|  
Db 41 QWNGSTWKQVEDH 53

## RESULT 2

US-10-793-626-2760  
; Sequence 2760, Application US/10793626  
; Publication No. US20050255478A1

## GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04

; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2760  
; LENGTH: 319  
; TYPE: PRT

## ORGANISM: Artificial Sequence

## FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-2760

Query Match 47.1%; Score 40; DB 6; Length 319;  
Best Local Similarity 63.6%; Pred. No. 20;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 WNGESEKPYDD 12  
:|||||:  
Db 46 FNKESEDPYDE 56

## RESULT 3

US-10-793-626-356  
; Sequence 356, Application US/10793626  
; Publication No. US20050255478A1

## GENERAL INFORMATION:

; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 356  
; LENGTH: 309  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-356

Query Match 45.9%; Score 39; DB 6; Length 309;  
Best Local Similarity 54.5%; Pred. No. 27;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 2 WNGESEKPYDD 12  
:|||||:  
Db 99 YRGESDLPFDD 109

## RESULT 4

US-11-024-959-484  
; Sequence 484, Application US/11024959  
; Publication No. US20060010516A1

## GENERAL INFORMATION:

; APPLICANT: FORSTER, RICHARD L.  
; APPLICANT: CONNETT, MARIE B.  
; APPLICANT: EMERSON, SARAH JANE  
; APPLICANT: GRIGOR, MURRAY ROBERT  
; APPLICANT: HIGGINS, COLLEEN M.  
; APPLICANT: LUND, STEVEN TROY  
; APPLICANT: MAGUSIN, ANDREAS  
; APPLICANT: KODRZYCKI, BOB  
; TITLE OF INVENTION: CELL CYCLE GENES AND RELATED METHODS  
; FILE REFERENCE: 044463-0360  
; CURRENT APPLICATION NUMBER: US/11/024,959  
; CURRENT FILING DATE: 2004-12-30  
; PRIOR APPLICATION NUMBER: 60/533,036  
; PRIOR FILING DATE: 2003-12-30

```
; ORGANISM: Pinus radiata
US-11-024-959-474

Query Match      44.1%; Score 37.5; DB 7; Length 488;
Best Local Similarity 40.0%; Pred. No. 75;
Matches 6; Conservative 5; Mismatches 1; Indels 3; Gaps 1;

Qy      2 WNGESEKPYDDH 13
Db      331 WNOQSQPVLKYNEH 345

RESULT 7
US-11-098-686-10175
; Sequence 10175, Application US/11098686
; Publication No. US2006024696A1
; GENERAL INFORMATION:
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES
; FILE REFERENCE: 09531-128001
; CURRENT APPLICATION NUMBER: US/11/098,686
; CURRENT FILING DATE: 2005-04-04
; PRIOR APPLICATION NUMBER: PCT/US03/31318
; PRIOR FILING DATE: 2003-10-01
; PRIOR APPLICATION NUMBER: US 60/416,395
; PRIOR FILING DATE: 2002-10-04
; NUMBER OF SEQ ID NOS: 11433
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10175
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Lawsonia intracellularis
US-11-098-686-10175

Query Match      43.5%; Score 37; DB 7; Length 168;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 6; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy      1 QMNGESEKPYDD 12
Db      39 RWKGEAKLSDD 50

RESULT 8
US-10-821-234-1479
; Sequence 1479, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes version 1.0
; SEQ ID NO 1479
; LENGTH: 748
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1479

Query Match      43.5%; Score 37; DB 6; Length 748;
Best Local Similarity 53.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy      2 WNGESEKPYDDHL 14
Db      168 RWNRGEAEDGHDH 181

RESULT 6
US-11-024-959-474
; Sequence 474, Application US/11024959
; Publication No. US20060010516A1
; GENERAL INFORMATION:
; APPLICANT: FORSTER, RICHARD L.
; APPLICANT: CONNETT, MARIE B.
; APPLICANT: EMERSON, SARAH JANE
; APPLICANT: GRIGOR, MURRAY ROBERT
; APPLICANT: HIGGINS, COLLEEN M.
; APPLICANT: LUND, STEVEN TROY
; APPLICANT: MAGUSIN, ANDREAS
; APPLICANT: KODRZYCKI, BOB
; TITLE OF INVENTION: CELL CYCLE GENES AND RELATED METHODS
; FILE REFERENCE: 044463-0360
; CURRENT APPLICATION NUMBER: US/11/024,959
; CURRENT FILING DATE: 2004-12-30
; PRIOR APPLICATION NUMBER: 60/533,036
; PRIOR FILING DATE: 2003-12-30
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 474
; LENGTH: 488
; TYPE: PRT
```

```
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 484
; LENGTH: 603
; TYPE: PRT
; ORGANISM: Pinus radiata
US-11-024-959-484

Query Match      45.3%; Score 38.5; DB 7; Length 603;
Best Local Similarity 47.1%; Pred. No. 64;
Matches 8; Conservative 2; Mismatches 4; Indels 3; Gaps 1;

Qy      1 QW---NGESEKPYDDHL 14
Db      426 QWDMNNGSEITQDYDQHL 442

RESULT 5
US-11-129-143-53
; Sequence 53, Application US/11129143
; Publication No. US20050266518A1
; GENERAL INFORMATION:
; APPLICANT: BERRY, Alan
; APPLICANT: BRETZEL, Werner
; APPLICANT: HUMBELIN, Markus
; APPLICANT: LOPEZ-ULIBARRI, Rual
; APPLICANT: MAYER, Anne F.
; APPLICANT: YELISEEV, Alexei A.
; TITLE OF INVENTION: IMPROVED ISOPRENOID PRODUCTION
; FILE REFERENCE: C38435/121966
; CURRENT APPLICATION NUMBER: US/11/129,143
; CURRENT FILING DATE: 2005-05-13
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 53
; LENGTH: 332
; TYPE: PRT
; ORGANISM: Paracoccus sp. R114
US-11-129-143-53

Query Match      44.1%; Score 37.5; DB 7; Length 332;
Best Local Similarity 50.0%; Pred. No. 51;
Matches 7; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

Qy      1 QWN-GESEKPYDDH 13
Db      168 RWNRGEAEDGHDH 181

RESULT 6
US-11-024-959-474
; Sequence 474, Application US/11024959
; Publication No. US20060010516A1
; GENERAL INFORMATION:
; APPLICANT: FORSTER, RICHARD L.
; APPLICANT: CONNETT, MARIE B.
; APPLICANT: EMERSON, SARAH JANE
; APPLICANT: GRIGOR, MURRAY ROBERT
; APPLICANT: HIGGINS, COLLEEN M.
; APPLICANT: LUND, STEVEN TROY
; APPLICANT: MAGUSIN, ANDREAS
; APPLICANT: KODRZYCKI, BOB
; TITLE OF INVENTION: CELL CYCLE GENES AND RELATED METHODS
; FILE REFERENCE: 044463-0360
; CURRENT APPLICATION NUMBER: US/11/024,959
; CURRENT FILING DATE: 2004-12-30
; PRIOR APPLICATION NUMBER: 60/533,036
; PRIOR FILING DATE: 2003-12-30
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 474
; LENGTH: 488
; TYPE: PRT
```

Db 230 WIPESENPDDDKI 242

## RESULT 9

US-10-453-372-170  
; Sequence 170, Application US/104533372  
; Publication No. US20060003323A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsbrook, et al.  
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
; FILE REFERENCE: 21402-589 A  
; CURRENT APPLICATION NUMBER: US/10/453,372  
; CURRENT FILING DATE: 2003-06-03  
; PRIOR APPLICATION NUMBER: 09/789390  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: 60/185967  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 09/823187  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195792  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR APPLICATION NUMBER: 09/839446  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/199476  
; PRIOR FILING DATE: 2000-03-25  
; PRIOR APPLICATION NUMBER: 09/863776  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: 60/208263  
; PRIOR FILING DATE: 2000-05-31  
; PRIOR APPLICATION NUMBER: 09/939398  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: 60/227800  
; PRIOR FILING DATE: 2000-08-25  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1609  
; SOFTWARE: CuraseqList version 0.1  
; SEQ ID NO 170  
; LENGTH: 2333  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-453-372-170

Query Match 43.5%; Score 37; DB 6; Length 2333;  
Best Local Similarity 63.6%; Pred. No. 4.2e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 NGESEKPYDDH 13

Db 1413 NIRTEKIYDDH 1423

## RESULT 10

US-10-453-372-114  
; Sequence 114, Application US/104533372  
; Publication No. US20060003323A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsbrook, et al.  
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
; FILE REFERENCE: 21402-589 A  
; CURRENT APPLICATION NUMBER: US/10/453,372  
; CURRENT FILING DATE: 2003-06-03  
; PRIOR APPLICATION NUMBER: 09/789390  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: 60/185967  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 09/823187  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195792  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR APPLICATION NUMBER: 09/839446  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/199476  
; PRIOR FILING DATE: 2000-03-25  
; Remaining Prior Application data removed - See File Wrapper or PALM.

; PRIOR APPLICATION NUMBER: 09/863776  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: 60/208263  
; PRIOR FILING DATE: 2000-05-31  
; PRIOR APPLICATION NUMBER: 09/939398  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: 60/227800  
; PRIOR FILING DATE: 2000-08-25  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1609  
; SOFTWARE: CuraseqList version 0.1  
; SEQ ID NO 114  
; LENGTH: 2662  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-453-372-114

Query Match 43.5%; Score 37; DB 6; Length 2662;  
Best Local Similarity 63.6%; Pred. No. 4.8e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 NGESEKPYDDH 13

Db 1742 NIRTEKIYDDH 1752

## RESULT 11

US-10-453-372-148  
; Sequence 148, Application US/104533372  
; Publication No. US20060003323A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsbrook, et al.  
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD  
; FILE REFERENCE: 21402-589 A  
; CURRENT APPLICATION NUMBER: US/10/453,372  
; CURRENT FILING DATE: 2003-06-03  
; PRIOR APPLICATION NUMBER: 09/789390  
; PRIOR FILING DATE: 2001-02-23  
; PRIOR APPLICATION NUMBER: 60/185967  
; PRIOR FILING DATE: 2000-03-01  
; PRIOR APPLICATION NUMBER: 09/823187  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195792  
; PRIOR FILING DATE: 2000-03-10  
; PRIOR APPLICATION NUMBER: 09/839446  
; PRIOR FILING DATE: 2001-03-19  
; PRIOR APPLICATION NUMBER: 60/199476  
; PRIOR FILING DATE: 2000-03-25  
; PRIOR APPLICATION NUMBER: 09/863776  
; PRIOR FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: 60/208263  
; PRIOR FILING DATE: 2000-05-31  
; PRIOR APPLICATION NUMBER: 09/939398  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: 60/227800  
; PRIOR FILING DATE: 2000-08-25  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 1609  
; SOFTWARE: CuraseqList version 0.1  
; SEQ ID NO 148  
; LENGTH: 2724  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-453-372-148

Query Match 43.5%; Score 37; DB 6; Length 2724;  
Best Local Similarity 63.6%; Pred. No. 4.9e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 NGESEKPYDDH 13

Db 1804 NIRTEKIYDDH 1814

```
RESULT 12
US-10-453-372-136
; Sequence 136, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: Curaseq1ist version 0.1
; SEQ ID NO 136
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-136

Query Match 43.5%; Score 37; DB 6; Length 2733;
Best Local Similarity 63.6%; Pred. No. 4.9e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 NGESEKPYDDH 13
Db 1813 NIRTEKIYDDH 1823

RESULT 13
US-10-453-372-142
; Sequence 142, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: Curaseq1ist version 0.1
; SEQ ID NO 136
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-136

Query Match 43.5%; Score 37; DB 6; Length 2733;
Best Local Similarity 63.6%; Pred. No. 4.9e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 NGESEKPYDDH 13
Db 1813 NIRTEKIYDDH 1823

RESULT 14
US-10-453-372-146
; Sequence 146, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: Curaseq1ist version 0.1
; SEQ ID NO 146
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-146

Query Match 43.5%; Score 37; DB 6; Length 2733;
Best Local Similarity 63.6%; Pred. No. 4.9e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 NGESEKPYDDH 13
Db 1813 NIRTEKIYDDH 1823

RESULT 15
```

```
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: Curaseq1ist version 0.1
; SEQ ID NO 142
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-142

Query Match 43.5%; Score 37; DB 6; Length 2733;
Best Local Similarity 63.6%; Pred. No. 4.9e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 NGESEKPYDDH 13
Db 1813 NIRTEKIYDDH 1823

RESULT 14
US-10-453-372-146
; Sequence 146, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: Curaseq1ist version 0.1
; SEQ ID NO 146
; LENGTH: 2733
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-146

Query Match 43.5%; Score 37; DB 6; Length 2733;
Best Local Similarity 63.6%; Pred. No. 4.9e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 NGESEKPYDDH 13
Db 1813 NIRTEKIYDDH 1823

RESULT 15
```

```
US-10-453-372-150
; Sequence 150, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 150
;   LENGTH: 2733
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-453-372-150
```

```
Query Match      43.5%; Score 37; DB 6; Length 2733;
Best Local Similarity 63.6%; Pred. No. 4.9e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      3 NGESEKPYDDH 13
      |  :||| |||
Db      1813 NIRTEKIYDDH 1823
```

```
RESULT 16
US-10-453-372-154
; Sequence 154, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
```

```
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 154
;   LENGTH: 2733
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-453-372-154
```

```
Query Match      43.5%; Score 37; DB 6; Length 2733;
Best Local Similarity 63.6%; Pred. No. 4.9e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      3 NGESEKPYDDH 13
      |  :||| |||
Db      1813 NIRTEKIYDDH 1823
```

```
RESULT 17
US-10-453-372-168
; Sequence 168, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 168
;   LENGTH: 2759
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-453-372-168
```

```
Query Match      43.5%; Score 37; DB 6; Length 2759;
Best Local Similarity 63.6%; Pred. No. 4.9e+02;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      3 NGESEKPYDDH 13
      |  :||| |||
Db      1839 NIRTEKIYDDH 1849
```

```
RESULT 18
US-10-453-372-116
; Sequence 116, Application US/10453372
```



```
; Publication No. US2006000323A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: Curaseqdist version 0.1
; SEQ ID NO 116
; LENGTH: 2765
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-116

Query Match      43.5%; Score 37; DB 6; Length 2765;
Best Local Similarity 63.6%; Pred. No. 4.9e+02;
Matches      7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      3 NGESEKPYDDH 13
      |::|::|::|
Db      1845 NIREKIYDDH 1855

RESULT 19
US-11-006-031-82
; Sequence 82, Application US/11006031
; Publication No. US20050272114A1
; GENERAL INFORMATION:
; APPLICANT: Wood, Keith V.
; APPLICANT: Klaubert, Dieter
; APPLICANT: Los, Georgyi V.
; APPLICANT: Bulleit, Robert F.
; APPLICANT: McDougall, Mark
; APPLICANT: Zimprich, Chad
; APPLICANT: Promega Corporation
; TITLE OF INVENTION: Substrates for Covalent Tethering to Proteins
; FILE REFERENCE: 341.032US1
; CURRENT APPLICATION NUMBER: US/11/006,031
; CURRENT FILING DATE: 2004-12-06
; PRIOR APPLICATION NUMBER: US 10/768,976
; PRIOR FILING DATE: 2004-01-30
; PRIOR APPLICATION NUMBER: US 60/444,094
; PRIOR FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: US 60/474,659
; PRIOR FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: US 60/592,499
; PRIOR FILING DATE: 2004-07-30
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 82
; LENGTH: 293
; TYPE: PRT

Query Match      43.5%; Score 37; DB 6; Length 2765;
Best Local Similarity 63.6%; Pred. No. 4.9e+02;
Matches      7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      3 NGESEKPYDDH 13
      |::|::|::|
Db      1845 NIREKIYDDH 1855

US-10-453-372-116
```

```
; TYPE: PRT
; ORGANISM: Rhodococcus rhodochrous
US-11-006-031-82

Query Match      42.9%; Score 36.5; DB 7; Length 293;
Best Local Similarity 43.8%; Pred. No. 65;
Matches      7; Conservative 4; Mismatches 0; Indels 5; Gaps 1;

Qy      4 GESEKP-----YDDHL 14
      |::|::|::|::|
Db      70 GKSDKPDLDYFFDDHV 85

RESULT 20
US-11-194-110-82
; Sequence 82, Application US/11194110
; Publication No. US20060024808A1
; GENERAL INFORMATION:
; APPLICANT: Darzins, Al
; APPLICANT: Encell, Lance
; APPLICANT: Klaubert, Dieter
; APPLICANT: Los, Georgyi V.
; APPLICANT: McDougall, Mark
; APPLICANT: Wood, Keith V.
; APPLICANT: Wood, Monika G.
; APPLICANT: Zimprich, Chad
; APPLICANT: Promega Corporation
; TITLE OF INVENTION: Covalent Tethering of Functional Groups to Proteins and Substrate
; FILE REFERENCE: 341.032US1
; CURRENT APPLICATION NUMBER: US/11/194,110
; CURRENT FILING DATE: 2005-07-29
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 82
; LENGTH: 293
; TYPE: PRT
; ORGANISM: Rhodococcus rhodochrous
US-11-194-110-82

Query Match      42.9%; Score 36.5; DB 7; Length 293;
Best Local Similarity 43.8%; Pred. No. 65;
Matches      7; Conservative 4; Mismatches 0; Indels 5; Gaps 1;

Qy      4 GESEKP-----YDDHL 14
      |::|::|::|::|
Db      70 GKSDKPDLDYFFDDHV 85

RESULT 21
US-11-194-110-48
; Sequence 48, Application US/11194110
; Publication No. US20060024808A1
; GENERAL INFORMATION:
; APPLICANT: Darzins, Al
; APPLICANT: Encell, Lance
; APPLICANT: Klaubert, Dieter
; APPLICANT: Los, Georgyi V.
; APPLICANT: McDougall, Mark
; APPLICANT: Wood, Keith V.
; APPLICANT: Wood, Monika G.
; APPLICANT: Zimprich, Chad
; APPLICANT: Promega Corporation
; TITLE OF INVENTION: Covalent Tethering of Functional Groups to Proteins and Substrate
; FILE REFERENCE: 341.032US1
; CURRENT APPLICATION NUMBER: US/11/194,110
; CURRENT FILING DATE: 2005-07-29
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 296
; TYPE: PRT
```

; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: A synthetic peptide  
US-11-194-110-48

Query Match 42.9%; Score 36.5; DB 7; Length 296;  
Best Local Similarity 43.8%; Pred. No. 66;  
Matches 7; Conservative 4; Mismatches 0; Indels 5; Gaps 1;

QY 4 GESEKP-----YDHL 14  
|:|:|:|:|:|:  
Db 71 GKSDKPDLDYFFDDHV 86

RESULT 22  
US-11-052-554A-70  
; Sequence 70, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 70  
; LENGTH: 248  
; TYPE: PRT  
; ORGANISM: Escherichia coli 0157:H7  
US-11-052-554A-70

Query Match 42.4%; Score 36; DB 7; Length 248;  
Best Local Similarity 50.0%; Pred. No. 67;  
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 QWNGSEKPY 10  
|:|:|:|:|:|:  
Db 163 QWNSQNEY 172

RESULT 23  
US-11-052-554A-342  
; Sequence 342, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 342  
; LENGTH: 248  
; TYPE: PRT  
; ORGANISM: Shigella flexneri 2a str. 2457T  
US-11-052-554A-342

Query Match 42.4%; Score 36; DB 7; Length 248;  
Best Local Similarity 50.0%; Pred. No. 67;  
Matches 5; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 QWNGSEKPY 10  
|:|:|:|:|:|:  
Db 163 QWNSQNEY 172

RESULT 24  
US-11-129-143-85  
; Sequence 85, Application US/11129143  
; Publication No. US20050266518A1  
; GENERAL INFORMATION:  
; APPLICANT: BERRY, Alan  
; APPLICANT: BRETZEL, Werner  
; APPLICANT: HUMBELIN, Markus  
; APPLICANT: LOPEZ-ULIBARRI, Rual  
; APPLICANT: MAYER, Anne F.  
; APPLICANT: YELISEEV, Alexei A.  
; TITLE OF INVENTION: IMPROVED ISOPRENOID PRODUCTION  
; FILE REFERENCE: C38435/121966  
; CURRENT FILING DATE: 2005-05-13  
; CURRENT FILING DATE: 2005-05-13  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 85  
; LENGTH: 317  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-11-129-143-85

Query Match 42.4%; Score 36; DB 7; Length 317;  
Best Local Similarity 53.8%; Pred. No. 85;  
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 2 WNGSEKPYDDHL 14  
|:|:|:|:|:|:  
Db 202 WVRQSEKYQDML 214

RESULT 25  
US-10-131-826A-512  
; Sequence 512, Application US/10131826A  
; Publication No. US20050245730A1  
; GENERAL INFORMATION:  
; APPLICANT: Baker, Kevin P.  
; APPLICANT: Beresini, Maureen  
; APPLICANT: DeForge, Laura  
; APPLICANT: Desnoyers, Luc  
; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gao, Wei-Qiang  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Sherwood, Steven  
; APPLICANT: Smith, Victoria  
; APPLICANT: Stewart, Timothy A.  
; APPLICANT: Tumas, Daniel  
; APPLICANT: Watanabe, Colin K  
; APPLICANT: Wood, William  
; APPLICANT: Zhang, Zemin  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P3330R1C128  
; CURRENT FILING DATE: 2002-04-24  
; PRIOR FILING DATE: 1997-06-18  
; PRIOR APPLICATION NUMBER: 60/049911  
; PRIOR FILING DATE: 1997-06-18  
; PRIOR APPLICATION NUMBER: 60/056974  
; PRIOR FILING DATE: 1997-08-26  
; PRIOR APPLICATION NUMBER: 60/059113  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059115  
; PRIOR FILING DATE: 1997-09-17  
; PRIOR APPLICATION NUMBER: 60/059117

```
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 512
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-826A-512

Query Match      42.4%; Score 36; DB 6; Length 716;
Best Local Similarity 60.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      2 WNGESEKPYD 11
Db      688 WEGDSEKDKD 697

RESULT 26
US-11-147-047-52
; Sequence 52, Application US/11147047
; Publication No. US20050260668A1
; GENERAL INFORMATION:
; APPLICANT: Agarwal, Pankaj
; APPLICANT: Murdock, Paul R.
; APPLICANT: Rizvi, Safia K.
; APPLICANT: Smith, Randall F.
; APPLICANT: Xiang, Zhaoqing
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50016
; CURRENT APPLICATION NUMBER: US/11/147,047
; CURRENT FILING DATE: 2005-06-07
; PRIOR APPLICATION NUMBER: US/10/221,097
; PRIOR FILING DATE: 2002-09-06
; PRIOR APPLICATION NUMBER: PCT/US01/07143
; PRIOR FILING DATE: 2001-03-05
; PRIOR APPLICATION NUMBER: 60/187,107
; PRIOR FILING DATE: 2000-03-06
; PRIOR APPLICATION NUMBER: 60/236,874
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/188,916
; PRIOR FILING DATE: 2000-03-13
; PRIOR APPLICATION NUMBER: 60/237,846
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 52
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-147-047-52

Query Match      42.4%; Score 36; DB 7; Length 716;
Best Local Similarity 60.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      2 WNGESEKPYD 11
Db      688 WEGDSEKDKD 697

RESULT 27
US-11-142-867-2
```

```
; Sequence 2, Application US/11142867
; Publication No. US20060002943A1
; GENERAL INFORMATION:
; APPLICANT: SMITH, VICTORIA
; TITLE OF INVENTION: Compositions and Methods for the Treatment of Tumor of
; FILE REFERENCE: PS110RIUS
; CURRENT APPLICATION NUMBER: US/11/142,867
; CURRENT FILING DATE: 2005-05-31
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/14941
; PRIOR FILING DATE: 2000-05-30
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US01/17800
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US01/19692
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 10/028,072
; PRIOR FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: US 10/128,690
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 10/128,694
; PRIOR FILING DATE: 2002-04-23
; PRIOR APPLICATION NUMBER: US 10/147,486
; PRIOR FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: US 10/991,287
; PRIOR FILING DATE: 2004-11-17
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-142-867-2

Query Match      42.4%; Score 36; DB 7; Length 716;
Best Local Similarity 60.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy      2 WNGESEKPYD 11
Db      688 WEGDSEKDKD 697

RESULT 28
US-11-074-176-348
; Sequence 348, Application US/11074176
; Publication No. US20050250135A1
; GENERAL INFORMATION:
; APPLICANT: Klaenhammer, Todd R.
; APPLICANT: Russell, William M.
; APPLICANT: Altermann, Eric
; APPLICANT: McAuliffe, Olivia
; APPLICANT: Peril, Andrea Azcarate
; TITLE OF INVENTION: Nucleic Acid Sequences Encoding
; FILE REFERENCE: 5051-694
; CURRENT APPLICATION NUMBER: US/11/074,176
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: 60/551,161
; PRIOR FILING DATE: 2004-03-08
; NUMBER OF SEQ ID NOS: 381
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 348
; LENGTH: 799
; TYPE: PRT
; ORGANISM: Lactobacillus acidophilus
US-11-074-176-348

Query Match      42.4%; Score 36; DB 7; Length 799;
Best Local Similarity 42.9%; Pred. No. 2.1e+02;
```

Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 QWGESEKPYDDHL 14

||| |::|

Db 446 QWPEVHEPNDEL 459

RESULT 29

US-11-074-176-172

; Sequence 172, Application US/11074176

; Publication No. US20050250135A1

; GENERAL INFORMATION:

; APPLICANT: Klaenhammer, Todd R.

; APPLICANT: Russell, William M.

; APPLICANT: Altermann, Eric

; APPLICANT: McAuliffe, Olivia

; APPLICANT: Peril, Andrea Arcarate

; TITLE OF INVENTION: Nucleic Acid Sequences Encoding

; FILE REFERENCE: 5051-694

; CURRENT APPLICATION NUMBER: US/11/074,176

; PRIOR FILING DATE: 2005-03-07

; PRIOR FILING DATE: 2004-03-08

; NUMBER OF SEQ ID NOS: 381

; SOFTWARE: Fast-SEQ for Windows Version 4.0

; SEQ ID NO 172

; LENGTH: 805

; TYPE: PRT

; ORGANISM: Lactobacillus acidophilus

US-11-074-176-172

Query Match

Best Local Similarity 42.4%; Score 36; DB 7; Length 805;

Matches 6; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 1 QWGESEKPYDDHL 14

||| |::|

Db 452 QWPEVHEPNDEL 465

RESULT 30

US-10-995-561-679

; Sequence 679, Application US/10995561

; Publication No. US20050272054A1

; GENERAL INFORMATION:

; APPLICANT: CARGILL, Michele et al.

; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH

; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF

; FILE REFERENCE: CL001559

; CURRENT APPLICATION NUMBER: US/10/995,561

; CURRENT FILING DATE: 2004-11-24

; NUMBER OF SEQ ID NOS: 85702

; SOFTWARE: Fast-SEQ for Windows Version 4.0

; SEQ ID NO 679

; LENGTH: 825

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-995-561-679

Query Match

Best Local Similarity 42.4%; Score 36; DB 6; Length 825;

Matches 4; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 QWGESEKPYDDHL 14

||| |::|

Db 220 KWHNSYREPPEQHL 233

RESULT 31

US-11-124-367A-469

; Sequence 469, Application US/11124367A

; Publication No. US20060024700A1

; GENERAL INFORMATION:

; APPLICANT: Michele Cargill

; APPLICANT: Hongjin Huang

; TITLE OF INVENTION: Genetic Polymorphisms Associated with

; FILE REFERENCE: CL001519.ORD

; CURRENT APPLICATION NUMBER: US/11/124,367A

; PRIOR FILING DATE: 2005-05-09

; PRIOR FILING DATE: 2004-05-07

; PRIOR FILING DATE: 2004-06-25

; PRIOR FILING DATE: 2004-08-09

; NUMBER OF SEQ ID NOS: 34460

; SOFTWARE: Fast-SEQ for Windows Version 4.0

; SEQ ID NO 469

; LENGTH: 825

; TYPE: PRT

; ORGANISM: Homo sapiens

US-11-124-367A-469

Query Match

Best Local Similarity 42.4%; Score 36; DB 7; Length 825;

Matches 4; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Qy 1 QWGESEKPYDDHL 14

||| |::|

Db 220 KWHNSYREPPEQHL 233

RESULT 32

US-10-821-234-1179

; Sequence 1179, Application US/10821234

; Publication No. US20050255114A1

; GENERAL INFORMATION:

; APPLICANT: Labat, Ivan

; APPLICANT: Stache-Crain, Birgit

; APPLICANT: Andarmani, Susan

; APPLICANT: Tang, Y. Tom

; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia

; FILE REFERENCE: 821A

; CURRENT APPLICATION NUMBER: US/10/821,234

; CURRENT FILING DATE: 2004-04-07

; PRIOR FILING DATE: 2003-04-07

; NUMBER OF SEQ ID NOS: 1704

; SOFTWARE: pt\_seq\_genes Version 1.0

; SEQ ID NO 1179

; LENGTH: 829

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-821-234-1179

Query Match

Best Local Similarity 42.4%; Score 36; DB 6; Length 829;

Matches 5; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 WNGESEKPYDDHL 14

||| |::|

Db 600 WNNQNKPNPDSL 612

RESULT 33

US-11-104-923A-5

; Sequence 5, Application US/11104923A

; Publication No. US20050260650A1

; GENERAL INFORMATION:

; APPLICANT: Castle, John C.

; APPLICANT: Garrett-Engle, Philip W.

; APPLICANT: Kan, Zhengyan

; APPLICANT: Armour, Christopher D.

; APPLICANT: Raymond, Christopher K.  
; TITLE OF INVENTION: ALTERNATIVELY SPLICED ISOFORMS OF  
; FILE REFERENCE: RS0218Y  
; CURRENT APPLICATION NUMBER: US/11/104,923A  
; PRIOR FILING DATE: 2005-04-12  
; PRIOR APPLICATION NUMBER: US 60/561,828  
; PRIOR FILING DATE: 2004-04-13  
; PRIOR APPLICATION NUMBER: US 60/564,261  
; PRIOR FILING DATE: 2004-04-21  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 852  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-104-923A-5

Query Match 42.4%; Score 36; DB 7; Length 852;  
Best Local Similarity 28.6%; Pred. No. 2.2e+02;  
Matches 4; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY 1 QWNGSEKPYDDHL 14  
Db 220 KWHNSYREPEQHL 233

RESULT 34  
US-11-113-424-54  
; Sequence 54, Application US/11/113,424  
; Publication No. US2005026073A1  
; GENERAL INFORMATION:  
; APPLICANT: Gangolli et al.  
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-225  
; CURRENT APPLICATION NUMBER: US/11/113,424  
; PRIOR FILING DATE: 2005-04-21  
; PRIOR APPLICATION NUMBER: 60/256,704  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 60/311,590  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/257,314  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/311,613  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/315,617  
; PRIOR FILING DATE: 2001-08-29  
; PRIOR APPLICATION NUMBER: 60/307,506  
; PRIOR FILING DATE: 2001-07-24  
; PRIOR APPLICATION NUMBER: 60/322,358  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: 60/294,075  
; PRIOR FILING DATE: 2001-05-29  
; PRIOR APPLICATION NUMBER: 60/288,153  
; PRIOR FILING DATE: 2001-05-02  
; NUMBER OF SEQ ID NOS: 190  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 54  
; LENGTH: 1045  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-113-424-54

Query Match 42.4%; Score 36; DB 7; Length 1045;  
Best Local Similarity 66.7%; Pred. No. 2.7e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 5 ESEKPYDDH 13  
Db 130 KTEKIYDDH 138

RESULT 35

US-11-124-367A-388  
; Sequence 388, Application US/11/124,367A  
; Publication No. US20060024700A1  
; GENERAL INFORMATION:  
; APPLICANT: Michele Cargill  
; APPLICANT: Hongjin Huang  
; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
; FILE REFERENCE: CL001519.ORD  
; CURRENT APPLICATION NUMBER: US/11/124,367A  
; CURRENT FILING DATE: 2005-05-09  
; PRIOR APPLICATION NUMBER: US 60/568,846  
; PRIOR FILING DATE: 2004-05-07  
; PRIOR APPLICATION NUMBER: US 60/582,609  
; PRIOR FILING DATE: 2004-06-25  
; PRIOR APPLICATION NUMBER: US 60/599,554  
; PRIOR FILING DATE: 2004-08-09  
; NUMBER OF SEQ ID NOS: 34460  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 388  
; LENGTH: 1047  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-124-367A-388

Query Match 42.4%; Score 36; DB 7; Length 1047;  
Best Local Similarity 44.4%; Pred. No. 2.8e+02;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 WNGSEKPY 10  
Db 265 WNSDTQNPY 273

RESULT 36  
US-11-124-367A-386  
; Sequence 386, Application US/11/124,367A  
; Publication No. US20060024700A1  
; GENERAL INFORMATION:  
; APPLICANT: Michele Cargill  
; APPLICANT: Hongjin Huang  
; TITLE OF INVENTION: Genetic Polymorphisms Associated with  
; FILE REFERENCE: CL001519.ORD  
; CURRENT APPLICATION NUMBER: US/11/124,367A  
; CURRENT FILING DATE: 2005-05-09  
; PRIOR APPLICATION NUMBER: US 60/568,846  
; PRIOR FILING DATE: 2004-05-07  
; PRIOR APPLICATION NUMBER: US 60/582,609  
; PRIOR FILING DATE: 2004-06-25  
; PRIOR APPLICATION NUMBER: US 60/599,554  
; PRIOR FILING DATE: 2004-08-09  
; NUMBER OF SEQ ID NOS: 34460  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 386  
; LENGTH: 1058  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-124-367A-386

Query Match 42.4%; Score 36; DB 7; Length 1058;  
Best Local Similarity 44.4%; Pred. No. 2.8e+02;  
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 WNGSEKPY 10  
Db 276 WNSDTQNPY 284

RESULT 37  
US-11-124-367A-387  
; Sequence 387, Application US/11/124,367A  
; Publication No. US20060024700A1

```
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; FILE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 387
; LENGTH: 1062
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-124-367A-387

Query Match      42.4%; Score 36; DB 7; Length 1062;
Best Local Similarity 44.4%; Pred. No. 2.8e+02;
Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY      2 WNGESEKPY 10
Db      280 WNSDTQNPY 288

RESULT 38
US-10-821-234-1097
; Sequence 1097, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1097
; LENGTH: 1094
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1097

Query Match      42.4%; Score 36; DB 6; Length 1094;
Best Local Similarity 66.7%; Pred. No. 2.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      5 ESEKPYDDH 13
Db      179 KTEKIYDDH 187

RESULT 39
US-11-096-051-4
; Sequence 4, Application US/11096051
; Publication No. US20050244868A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: MacLachlan, Timothy K
; APPLICANT: Rastelli, Luca
; APPLICANT: Vernet, Corine
; APPLICANT: Ettenberg, Seth
```

```
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use
; FILE REFERENCE: Attorney Docket No. Cura 967
; CURRENT APPLICATION NUMBER: US/11/096,051
; CURRENT FILING DATE: 2005-03-30
; PRIOR APPLICATION NUMBER: 10/038,854
; PRIOR FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 10/455,772
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: 60/557,978
; PRIOR FILING DATE: 2004-03-30
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 4
; LENGTH: 2376
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-096-051-4

Query Match      42.4%; Score 36; DB 7; Length 2376;
Best Local Similarity 66.7%; Pred. No. 6.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      5 ESEKPYDDH 13
Db      1461 KTEKIYDDH 1469

RESULT 40
US-11-096-051-2
; Sequence 2, Application US/11096051
; Publication No. US20050244868A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: MacLachlan, Timothy K
; APPLICANT: Rastelli, Luca
; APPLICANT: Vernet, Corine
; APPLICANT: Ettenberg, Seth
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use
; FILE REFERENCE: Attorney Docket No. Cura 967
; CURRENT APPLICATION NUMBER: US/11/096,051
; CURRENT FILING DATE: 2005-03-30
; PRIOR APPLICATION NUMBER: 10/038,854
; PRIOR FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 10/455,772
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: 60/557,978
; PRIOR FILING DATE: 2004-03-30
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 2
; LENGTH: 2715
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-096-051-2

Query Match      42.4%; Score 36; DB 7; Length 2715;
Best Local Similarity 66.7%; Pred. No. 7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      5 ESEKPYDDH 13
Db      1800 KTEKIYDDH 1808

RESULT 41
US-11-113-424-51
; Sequence 51, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
```

; CURRENT FILING DATE: 2005-04-21  
; PRIOR APPLICATION NUMBER: 60/256,704  
; PRIOR FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 60/311,590  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/257,314  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/311,613  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: 60/315,617  
; PRIOR FILING DATE: 2001-08-29  
; PRIOR APPLICATION NUMBER: 60/307,506  
; PRIOR FILING DATE: 2001-07-24  
; PRIOR APPLICATION NUMBER: 60/322,358  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: 60/294,075  
; PRIOR FILING DATE: 2001-05-29  
; PRIOR APPLICATION NUMBER: 60/288,153  
; PRIOR FILING DATE: 2001-05-02  
; NUMBER OF SEQ ID NOS: 190  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 51  
; LENGTH: 2715  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-11-113-424-51

Query Match 42.4%; Score 36; DB 7; Length 2715;  
Best Local Similarity 66.7%; Pred. No. 7e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ESEKPYDDH 13  
Db 1800 KTEKIYDDH 1808

RESULT 42  
US-11-096-051-10  
; Sequence 10, Application US/11096051  
; Publication No. US20050244868A1  
; GENERAL INFORMATION:  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: MacLachlan, Timothy K  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Vernet, Corine  
; APPLICANT: Ettenberg, Seth  
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use  
; FILE REFERENCE: Attorney Docket No. Cura 967  
; CURRENT APPLICATION NUMBER: US/11/096,051  
; CURRENT FILING DATE: 2005-03-30  
; PRIOR APPLICATION NUMBER: 10/038,854  
; PRIOR FILING DATE: 2001-12-31  
; PRIOR APPLICATION NUMBER: 10/455,772  
; PRIOR FILING DATE: 2003-06-04  
; PRIOR APPLICATION NUMBER: 60/557,978  
; PRIOR FILING DATE: 2004-03-30  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: CuraSeqList version 0.1  
; SEQ ID NO 10  
; LENGTH: 2721  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-096-051-10

Query Match 42.4%; Score 36; DB 7; Length 2721;  
Best Local Similarity 66.7%; Pred. No. 7e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ESEKPYDDH 13  
Db 1806 KTEKIYDDH 1814

RESULT 43  
US-11-096-051-8  
; Sequence 8, Application US/11096051  
; Publication No. US20050244868A1  
; GENERAL INFORMATION:  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: MacLachlan, Timothy K  
; APPLICANT: Rastelli, Luca  
; APPLICANT: Vernet, Corine  
; APPLICANT: Ettenberg, Seth  
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use  
; FILE REFERENCE: Attorney Docket No. Cura 967  
; CURRENT APPLICATION NUMBER: US/11/096,051  
; CURRENT FILING DATE: 2005-03-30  
; PRIOR APPLICATION NUMBER: 10/038,854  
; PRIOR FILING DATE: 2001-12-31  
; PRIOR APPLICATION NUMBER: 10/455,772  
; PRIOR FILING DATE: 2003-06-04  
; PRIOR APPLICATION NUMBER: 60/557,978  
; PRIOR FILING DATE: 2004-03-30  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: CuraSeqList version 0.1  
; SEQ ID NO 8  
; LENGTH: 2725  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-096-051-8

Query Match 42.4%; Score 36; DB 7; Length 2725;  
Best Local Similarity 66.7%; Pred. No. 7e+02;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ESEKPYDDH 13  
Db 1810 KTEKIYDDH 1818

RESULT 44  
US-11-024-959-371  
; Sequence 371, Application US/11024959  
; Publication No. US20060010516A1  
; GENERAL INFORMATION:  
; APPLICANT: FORSTER, RICHARD L.  
; APPLICANT: CONNETT, MARIE B.  
; APPLICANT: EMERSON, SARAH JANE  
; APPLICANT: GRIGOR, MURRAY ROBERT  
; APPLICANT: HIGGINS, COLLEEN M.  
; APPLICANT: LUND, STEVEN TROY  
; APPLICANT: MAGUSIN, ANDREAS  
; APPLICANT: KODRZYCKI, BOB  
; TITLE OF INVENTION: CELL CYCLE GENES AND RELATED METHODS  
; FILE REFERENCE: 044463-0360  
; CURRENT APPLICATION NUMBER: US/11/024,959  
; CURRENT FILING DATE: 2004-12-30  
; PRIOR APPLICATION NUMBER: 60/533,036  
; PRIOR FILING DATE: 2003-12-30  
; NUMBER OF SEQ ID NOS: 782  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 371  
; LENGTH: 523  
; TYPE: PRT  
; ORGANISM: Eucalyptus sp.  
US-11-024-959-371

Query Match 41.8%; Score 35.5; DB 7; Length 523;  
Best Local Similarity 41.2%; Pred. No. 1.7e+02;  
Matches 7; Conservative 3; Mismatches 4; Indels 3; Gaps 1;

Qy 1 QWN---GESEKPYDDHL 14  
Db 346 QWDMNTGQITQYDQHL 362

RESULT 45  
US-10-467-657-2762  
; Sequence 2762, Application US/10467657  
; Publication No. US20050260581A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; APPLICANT: FONTANA Maria Rita  
; APPLICANT: PIZZA Mariagrazia  
; APPLICANT: MASIGNANI Vega  
; APPLICANT: MONACI Elisabetta  
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/467,657  
; CURRENT FILING DATE: 2003-08-11  
; PRIOR APPLICATION NUMBER: GB-0103424.8  
; PRIOR FILING DATE: 2001-02-12  
; NUMBER OF SEQ ID NOS: 9218  
; SOFTWARE: SeqWin99, version 1.04  
; SEQ ID NO 2762  
; LENGTH: 82  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-2762

Query Match 41.2%; Score 35; DB 6; Length 82;  
Best Local Similarity 66.7%; Pred. No. 33;  
Matches 6; Conservative 0; Mismatches 3; Indels 3; Gaps 0;

QY 4 GESEKPYDD 12  
| | | | |  
DB 64 GAKEPPYDD 72

RESULT 46  
US-11-054-515-1513  
; Sequence 1513, Application US/11054515  
; Publication No. US2005025532A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruben et al.  
; TITLE OF INVENTION: Antibodies that Immunospecifically Bind Blys  
; FILE REFERENCE: PF523P3  
; CURRENT APPLICATION NUMBER: US/11/054,515  
; CURRENT FILING DATE: 2005-02-10  
; PRIOR APPLICATION NUMBER: 60/543,296  
; PRIOR FILING DATE: 2004-02-11  
; PRIOR APPLICATION NUMBER: 60/580,347  
; PRIOR FILING DATE: 2004-06-18  
; PRIOR APPLICATION NUMBER: 10/293,418  
; PRIOR FILING DATE: 2002-11-14  
; PRIOR APPLICATION NUMBER: 60/331,469  
; PRIOR FILING DATE: 2001-11-16  
; PRIOR APPLICATION NUMBER: 60/340,817  
; PRIOR FILING DATE: 2001-12-19  
; PRIOR APPLICATION NUMBER: 09/880,748  
; PRIOR FILING DATE: 2001-06-15  
; PRIOR APPLICATION NUMBER: 60/293,499  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: 60/277,379  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: 60/276,248  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/240,816  
; PRIOR FILING DATE: 2000-10-17  
; Remaining Prior Application data removed - See File Wrapper or PALM.  
; NUMBER OF SEQ ID NOS: 3247  
; SEQ ID NO 1513  
; LENGTH: 244  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-054-515-1513  
  
Query Match 41.2%; Score 35; DB 7; Length 244;  
Best Local Similarity 41.7%; Pred. No. 95;

Matches 5; Conservative 3; Mismatches 4; Indels 0; Gaps 0;  
QY 1 QWNGSEKPYDD 12  
| | | | |  
DB 46 EWMGGSDPPYGE 57

RESULT 47  
US-11-156-084-215  
; Sequence 215, Application US/11156084  
; Publication No. US20060010515A1  
; GENERAL INFORMATION:  
; APPLICANT: Monsanto Technology LLC  
; TITLE OF INVENTION: Controlled expression of cytokinin biosynthetic genes leads to agronomically interesting phenotypes  
; FILE REFERENCE: (38-21)  
; CURRENT APPLICATION NUMBER: US/11/156,084  
; CURRENT FILING DATE: 2005-06-17  
; NUMBER OF SEQ ID NOS: 364  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 215  
; LENGTH: 305  
; TYPE: PRT  
; ORGANISM: Listeria innocua  
US-11-156-084-215

Query Match 41.2%; Score 35; DB 7; Length 305;  
Best Local Similarity 60.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 5 ESEKPYDDHL 14  
| | | | |  
DB 219 EASKLYDEHL 228

RESULT 48  
US-11-098-686-10238  
; Sequence 10238, Application US/11098686  
; Publication No. US20060024696A1  
; GENERAL INFORMATION:  
; APPLICANT: Kapur, Vivek and Gebhart, Connie J.  
; TITLE OF INVENTION: NUCLEIC ACID AND POLYPEPTIDE SEQUENCES  
; FILE REFERENCE: 09531-128001  
; CURRENT APPLICATION NUMBER: US/11/098,686  
; CURRENT FILING DATE: 2005-04-04  
; PRIOR APPLICATION NUMBER: PCT/US03/31318  
; PRIOR FILING DATE: 2003-10-01  
; PRIOR APPLICATION NUMBER: US 60/416,395  
; PRIOR FILING DATE: 2002-10-04  
; NUMBER OF SEQ ID NOS: 11433  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10238  
; LENGTH: 350  
; TYPE: PRT  
; ORGANISM: Lawsonia intracellularis  
US-11-098-686-10238

Query Match 41.2%; Score 35; DB 7; Length 350;  
Best Local Similarity 54.5%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 WNGSEKPYDD 12  
| | | | |  
DB 273 WNFAPKPYDD 283

RESULT 49  
US-10-063-703-32  
; Sequence 32, Application US/10063703  
; Publication No. US20060008901A1  
; GENERAL INFORMATION:  
; APPLICANT: Eaton, Dan L.



; APPLICANT: Filvaroff, Ellen  
; APPLICANT: Gerritsen, Mary E.  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Watanabe, Colin K.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
; FILE REFERENCE: P323OR1C1  
; CURRENT APPLICATION NUMBER: US/10/063,703  
; CURRENT FILING DATE: 2002-05-08  
; Prior Application removed - See Palm or File Wrapper  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 32  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
; US-10-063-703--32

Query Match 41.2%; Score 35; DB 6; Length 445;  
Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NGESEKPY 10  
; | | | | |  
Db 397 DGSSEKPY 404

RESULT 50  
US-11-102-240-32  
; Sequence 32, Application US/11102240  
; Publication No. US20050260647A1  
; GENERAL INFORMATION:  
; APPLICANT: Goddard, Audrey  
; APPLICANT: Godowski, Paul J.  
; APPLICANT: Grimaldi, Christopher J.  
; APPLICANT: Gurney, Austin L.  
; APPLICANT: Wood, William I.  
; TITLE OF INVENTION: ANTIBODIES TO POLYPEPTIDES ENCODED BY A NUCLEIC ACID UNDEREXPRESS  
; FILE REFERENCE: P323OR1C106C  
; CURRENT APPLICATION NUMBER: US/11/102,240  
; CURRENT FILING DATE: 2005-04-08  
; PRIOR APPLICATION NUMBER: 10/063662  
; PRIOR FILING DATE: 2002-05-07  
; PRIOR APPLICATION NUMBER: 10/006867  
; PRIOR FILING DATE: 2001-12-06  
; PRIOR APPLICATION NUMBER: PCT/US00/23328  
; PRIOR FILING DATE: 2000-08-24  
; PRIOR APPLICATION NUMBER: 60/170262  
; PRIOR FILING DATE: 199-12-09  
; NUMBER OF SEQ ID NOS: 170  
; SEQ ID NO 32  
; LENGTH: 445  
; TYPE: PRT  
; ORGANISM: Homo Sapien  
; US-11-102-240-32

Query Match 41.2%; Score 35; DB 7; Length 445;  
Best Local Similarity 75.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NGESEKPY 10  
; | | | | |  
Db 397 DGSSEKPY 404

Search completed: February 10, 2006, 23:50:49  
Job time : 30 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:05:33 ; Search time 311.954 Seconds  
(without alignments)  
607.053 Million cell updates/sec

Title: US-10-797-821-29  
Perfect score: 2104  
Sequence: 1 MKRILSAVLVSGVTLSSAT.....SIGNRCWFNPGSVSYIYPN 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_21.\*  
1: Geneseq1980s.\*  
2: Geneseq1990s.\*  
3: Geneseq2000s.\*  
4: Geneseq2001s.\*  
5: Geneseq2002s.\*  
6: Geneseq2003as.\*  
7: Geneseq2003bs.\*  
8: Geneseq2004s.\*  
9: Geneseq2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2104	100.0	431	7	ADD93649 Streptococ
2	2104	100.0	431	9	ADX37272 Streptococ
3	2100	99.8	431	7	ADD93650 Streptococ
4	2100	99.8	431	9	ADX37273 Streptococ
5	2096	99.6	431	9	ABE91500 Microbial
6	2092	99.4	431	7	ADD93653 Streptococ
7	2092	99.4	431	9	ADX37276 Streptococ
8	2079.5	98.8	432	7	ADD93651 Streptococ
9	2079.5	98.8	432	9	ADX37274 Streptococ
10	2070.5	98.4	432	7	ADD93652 Streptococ
11	2070.5	98.4	432	9	ADX37275 Streptococ
12	1231	58.5	447	5	ABP29684 Streptococ
13	1231	58.5	447	8	ADU69524 S agalact
14	1231	58.5	447	8	ADV88392 Streptococ
15	1231	58.5	447	8	ADV81808 Streptococ
16	1231	58.5	447	8	ADV79645 Streptococ
17	1091.5	51.9	398	9	ABE91542 Microbial
18	1085.5	51.6	398	5	ABP25919 Streptococ
19	1085.5	51.6	398	8	ADR83884 S. pyogen
20	1073	51.0	395	5	ABP25918 Streptococ
21	1057.5	50.3	392	6	ABU02747 S. pneumo
22	1057.5	50.3	392	8	ADT50226 S_pneumon
23	1057.5	50.3	392	8	ADT50165 S_pneumon
24	1057.5	50.3	399	8	ADR94595 Novel S.

25	1057.5	50.3	399	9	AEA58465 Streptococ
26	1056.5	50.2	392	8	ADK47859 Streptococ
27	1056.5	50.2	392	8	ADT50227 S_pneumon
28	670	31.8	461	2	AAR14150 MSP encod
29	669	31.8	461	2	AAR14530 Usp45 pro
30	658.5	31.3	456	5	ABB55584 Lactococcc
31	591	28.1	524	6	ABU29734 Protein e
32	587	27.9	525	7	ADC95468 E. faeciu
33	562	26.7	210	2	AAY22579 Bacterial
34	524.5	24.9	482	9	ADV16553 E. faecal
35	524.5	24.9	497	7	ADH88105 Enterococ
36	507	24.1	449	2	AAV00250 Enterococ
37	507	24.1	449	5	ABP43469 E faecali
38	507	24.1	449	6	ABU88497 E. faecal
39	507	24.1	449	6	ABU13748 Enterococ
40	507	24.1	449	9	ADV16734 E. faecal
41	507	24.1	449	9	ADY39216 Novel Ent
42	507	24.1	450	7	ADH87829 Enterococ
43	458	21.8	422	2	RAY00251 Enterococ
44	458	21.8	422	5	ABP43470 E faecali
45	458	21.8	422	6	ABU88498 E. faecal

ALIGNMENTS

RESULT 1					
ADD93649					
ID	ADD93649 standard; protein; 431 AA.				
XX					
AC	ADD93649;				
XX					
DT	29-JAN-2004 (first entry)				
XX					
DE	Streptococcus mutans glucan binding protein-B.				
XX					
KW	Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.				
XX					
OS	Streptococcus mutans.				
XX					
FH	Key	Location/Qualifiers			
FT	Region	6..25	/note= "HLA-binding peptide"		
FT	Region	16..35	/note= "HLA-binding peptide"		
FT	Region	33..52	/note= "HLA-binding peptide"		
FT	Region	37..56	/note= "HLA-binding peptide"		
FT	Region	48..67	/note= "HLA-binding peptide"		
FT	Region	52..71	/note= "HLA-binding peptide"		
FT	Region	88..107	/note= "HLA-binding peptide"		
FT	Region	113..132	/note= "HLA-binding peptide"		
FT	Region	117..136	/note= "HLA-binding peptide"		
FT	Region	137..156	/note= "HLA-binding peptide"		
FT	Region	174..193	/note= "HLA-binding peptide"		
FT	Region	194..213	/note= "HLA-binding peptide"		
FT	Region	214..233	/note= "HLA-binding peptide"		
FT	Region	248..267	/note= "HLA-binding peptide"		
FT	Region	289..308	/note= "HLA-binding peptide"		
FT	Region	306..325	/note= "HLA-binding peptide"		
FT	Region		/note= "HLA-binding peptide"		

FT Region 311..330  
FT /note= "HLA-binding peptide"  
FT Region 349..368  
FT /note= "HLA-binding peptide"  
FT Region 365..384  
FT /note= "HLA-binding peptide"  
FT Region 383..402  
FT /note= "HLA-binding peptide"  
FT Region 403..422  
FT /note= "HLA-binding peptide"  
XX  
PN WO2003075845-A2.  
XX  
XX 18-SEP-2003.  
XX  
XX 07-MAR-2003; 2003WO-US006962.  
XX  
XX 07-MAR-2002; 2002US-0363209P.  
XX 08-AUG-2002; 2002US-0402483P.  
XX (FORS-) FORSYTH INST.  
XX  
XX Smith DJ, Taubman MA;  
XX  
XX WPI; 2003-845091/78.  
XX GENBANK; AY046410.  
XX  
XX Composition useful as vaccines for dental caries comprises a fragment of  
PT a glucan binding protein-B binding to a major histocompatibility complex  
PT class II protein.  
XX  
XX Claim 6; Page 7; 49pp; English.  
XX  
XX The present sequence is the protein sequence of the glucan binding  
CC protein-B (GbpB) of Streptococcus mutans strain SK32. The sequence  
CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
CC invention provides immunogenic compositions and vaccines for dental  
CC caries. The compositions comprise major histocompatibility complex (MHC)  
CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
CC covalently linked with peptide subunits of a glucosyltransferase. The  
CC compositions are used in a claimed method of eliciting production of an  
CC antibody in a mammal. Dipitopic or multiepitopic polypeptides can be  
CC prepared synthetically or by recombinant DNA technology. Antibodies  
CC raised against MHC class II binding fragments of GbpB can be used in  
CC passive immunisation.  
XX  
XX Sequence 431 AA;  
SQ  
Query Match 100.0%; Score 2104; DB 7; Length 431;  
Best Local Similarity 100.0%; Pred. No. 1.7e-134;  
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60  
Db 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60  
QY 61 QVSALQTOQAELOAENQRLAQAQATLGGQIQTLSKIVARNESLKQAARSQAQVNTIQ 120  
Db 61 QVSALQTOQAELOAENQRLAQAQATLGGQIQTLSKIVARNESLKQAARSQAQVNTIQ 120  
QY 121 INAININSKVSDAINRVSAIRREVVSANEKMLQOEODKAAVEQKQENQAINTVAANOE 180  
Db 121 INAININSKVSDAINRVSAIRREVVSANEKMLQOEODKAAVEQKQENQAINTVAANOE 180  
QY 181 TIAQNTNALNTQAQLEAAQLNLQAEITTAQDQKATLVAQKAAAEBAARQAAAAA 240  
Db 181 TIAQNTNALNTQAQLEAAQLNLQAEITTAQDQKATLVAQKAAAEBAARQAAAAA 240  
QY 241 KAAAEAKALQEAQAQVAQVANNNTQATDASDQAAADNTQAAQTGDSQSAQAQVNN 300  
Db 241 KAAAEAKALQEAQAQVAQVANNNTQATDASDQAAADNTQAAQTGDSQSAQAQVNN 300  
QY 301 DQESTTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNYWNGGQWAA 360

Db 301 DQESTTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNYWNGGQWAA 360  
QY 361 SAAAAGYRVGSTPSAGAVAVWMDGGYGHVAYVTGVOGQIQVQEAANYAGNOSTGNVRGW 420  
Db 361 SAAAAGYRVGSTPSAGAVAVWMDGGYGHVAYVTGVOGQIQVQEAANYAGNOSTGNVRGW 420  
QY 421 NPGSVSYIYPN 431  
Db 421 NPGSVSYIYPN 431  
RESULT 2  
AD37272  
ID AD37272 standard; protein; 431 AA.  
XX  
AC AD37272;  
XX  
DT 21-APR-2005 (first entry)  
XX  
DE Streptococcus mutant glucan binding protein B variant #1.  
XX  
KW immunogenicity; immune stimulation; glucan binding protein-B;  
KW microparticle; major histocompatibility complex; tooth disease.  
XX  
OS Streptococcus mutans.  
XX  
PN US2005031633-A1.  
XX  
PD 10-FEB-2005.  
XX  
PF 09-MAR-2004; 2004US-00797821.  
XX  
PR 13-APR-1998; 98US-0081550P.  
PR 08-JAN-1999; 99US-0115142P.  
PR 12-APR-1999; 99US-00290049.  
PR 07-MAR-2002; 2002US-0363209P.  
PR 08-AUG-2002; 2002US-0402483P.  
PR 07-MAR-2003; 2003US-00383930.  
XX  
PA (SMIT/) SMITH D J.  
PA (TAUB/) TAUBMAN M A.  
XX  
PI Smith DJ, Taubman MA;  
XX  
XX WPI; 2005-151644/16.  
XX  
XX New composition comprising a fragment of a glucan binding protein-B  
PT (GbpB) that binds to MHC class II protein, and a biocompatible  
PT microparticle, useful for producing an antibody (claimed) for immunizing  
PT mammals against dental caries.  
XX  
XX Claim 3; SEQ ID NO 29; 73pp; English.  
PS  
XX The invention relates to a composition comprising a fragment of a glucan  
CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
CC fragment binds to a major histocompatibility complex (MHC) class II  
CC protein. The composition is useful for producing an antibody for  
CC immunizing mammals against dental caries. This sequence corresponds to a  
CC Streptococcus mutans GbpB protein of the invention.  
XX  
XX Sequence 431 AA;  
SQ  
Query Match 100.0%; Score 2104; DB 9; Length 431;  
Best Local Similarity 100.0%; Pred. No. 1.7e-134;  
Matches 431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60  
Db 1 MKKRILSAVLVSGVTLSSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60  
QY 61 QVSALQTOQAELOAENQRLAQAQATLGGQIQTLSKIVARNESLKQAARSQAQVNTIQ 120

```
Db 61 QVSALQTQQAELQAEQNRLEAQSATLGGQQTQLSSKIVARNESLKQARSAQKSNATSY 120
Qy 121 INAIINSKVSVDAINRVSAIREVVSANEKMLQQOEODKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKVSVDAINRVSAIREVVSANEKMLQQOEODKAAVEQKQENQAAINTVAANQE 180
Qy 181 TIAQNTNALNTQQAELQAEQNRLEAQSATLGGQQTQLSSKIVARNESLKQARSAQKSNATSY 240
Db 181 TIAQNTNALNTQQAELQAEQNRLEAQSATLGGQQTQLSSKIVARNESLKQARSAQKSNATSY 240
Qy 241 KAAAEAKALQEQAAQAAQVAAANNNTQATDASDQQAADNTQAAQTGDSSTEQSAAQAVNNS 300
Db 241 KAAAEAKALQEQAAQAAQVAAANNNTQATDASDQQAADNTQAAQTGDSSTEQSAAQAVNNS 300
Qy 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGOWAA 360
Db 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGOWAA 360
Qy 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYVTGVGGQIQVQEAANYAGNQSIGNYRGWF 420
Db 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYVTGVGGQIQVQEAANYAGNQSIGNYRGWF 420
Qy 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431

RESULT 3
ADD93650
ID ADD93650 standard; protein; 431 AA.
XX
AC ADD93650;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucan binding protein-B.
XX
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus mutans.
XX
PN W02003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
(PORS-) FORSYTH INST.
XX
PA Smith DJ, Taubman MA;
PI WPI; 2003-845091/78.
XX GENBANK; AY046411.
DR
XX
Composition useful as vaccines for dental caries comprises a fragment of
a glucan binding protein-B binding to a major histocompatibility complex
class II protein.
XX
Claim 5; Page 8; 49pp; English.
XX
The present sequence is the protein sequence of the glucan binding
protein-B (GbpB) of Streptococcus mutans strain 3VF4. The sequence
includes a number of human leucocyte antigen (HLA)-binding epitopes. The
invention provides immunogenic compositions and vaccines for dental
caries. The compositions comprise major histocompatibility complex (MHC)
class II protein-binding GbpB peptides, especially HLA-binding peptides,
covalently linked with peptide subunits of a glucosyltransferase. The
compositions are used in a claimed method of eliciting production of an
antibody in a mammal. Diepitopic or multiepitopic polypeptides can be
prepared synthetically or by recombinant DNA technology. Antibodies
```

```
CC raised against MHC class II binding fragments of GbpB can be used in
CC passive immunisation.
SQ Sequence 431 AA;
Query Match 99.8%; Score 2100; DB 7; Length 431;
Best Local Similarity 99.8%; Pred No. 3.2e-134; Indels 0; Gaps 0;
Matches 430; Conservative 0; Mismatches 1;
Qy 1 MKKRILSAVLVSGVTLSSTATLTSVAVKADDFDAIQSDSKINNLTAQQAQAAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSSTATLTSVAVKADDFDAIQSDSKINNLTAQQAQAAQVNTIQ 60
Qy 61 QVSALQTQQAELQAEQNRLEAQSATLGGQQTQLSSKIVARNESLKQARSAQKSNATSY 120
Db 61 QVSALQTQQAELQAEQNRLEAQSATLGGQQTQLSSKIVARNESLKQARSAQKSNATSY 120
Qy 121 INAIINSKVSVDAINRVSAIREVVSANEKMLQQOEODKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKVSVDAINRVSAIREVVSANEKMLQQOEODKAAVEQKQENQAAINTVAANQE 180
Qy 181 TIAQNTNALNTQQAELQAEQNRLEAQSATLGGQQTQLSSKIVARNESLKQARSAQKSNATSY 240
Db 181 TIAQNTNALNTQQAELQAEQNRLEAQSATLGGQQTQLSSKIVARNESLKQARSAQKSNATSY 240
Qy 241 KAAAEAKALQEQAAQAAQVAAANNNTQATDASDQQAADNTQAAQTGDSSTEQSAAQAVNNS 300
Db 241 KAAAEAKALQEQAAQAAQVAAANNNTQATDASDQQAADNTQAAQTGDSSTEQSAAQAVNNS 300
Qy 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGOWAA 360
Db 301 DQESTTATEAQPSSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVGNWNGGOWAA 360
Qy 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYVTGVGGQIQVQEAANYAGNQSIGNYRGWF 420
Db 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAVYVTGVGGQIQVQEAANYAGNQSIGNYRGWF 420
Qy 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431

RESULT 4
ADD937273
ID ADD937273 standard; protein; 431 AA.
XX
AC ADD937273;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B variant #2.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
(SMIT/) SMITH D J.
(PAUB/) TAUBMAN M A.
PA
XX
```



```
Db 241 KAAAEAKALOEQAQAQAANNNTQATDVSDQQAADNTQAAQTGDSSTQSAQAQVNS 300
Qy 301 DQSTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360
Db 301 DQSTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360
Qy 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAYVTGVGGQIQVQEAANYAGNQSIGNYRGWF 420
Db 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAYVTGVGGQIQVQEAANYAGNQSIGNYRGWF 420
Qy 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431

RESULT 6
ADD93653
ID ADD93653 standard; protein; 431 AA.
XX
AC ADD93653;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucan binding protein-B.
XX
KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus mutans.
XX
PN W02003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
PA (FORS-) FORSYTH INST.
XX
PI Smith DJ, Taubman MA;
XX
WFI; 2003-845091/78.
DR GENBANK; AY046414.
XX
Composition useful as vaccines for dental caries comprises a fragment of
a glucan binding protein-B binding to a major histocompatibility complex
class II protein.
XX
Claim 5; Page 8-9; 49pp; English.
XX
The present sequence is the protein sequence of the glucan binding
protein-B (GbpB) of Streptococcus mutans strain 5SM3. The sequence
includes a number of human leucocyte antigen (HLA)-binding epitopes. The
invention provides immunogenic compositions and vaccines for dental
caries. The compositions comprise major histocompatibility complex (MHC)
class II protein-binding GbpB peptides, especially HLA-binding peptides,
covalently linked with peptide subunits of a glucosyltransferase. The
compositions are used in a claimed method of eliciting production of an
antibody in a mammal. Dipitopic or multiepitopic polypeptides can be
prepared synthetically or by recombinant DNA technology. Antibodies
raised against MHC class II binding fragments of GbpB can be used in
passive immunisation.
XX
Sequence 431 AA;
XX
Query Match 99.4%; Score 2092; DB 7; Length 431;
Best Local Similarity 99.3%; Pred. No. 1.le-133;
Matches 428; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 MKKRILSAVLVSGVTLSSTATLSAVKADDFDAQIASQDSKINNLTAAQQQAQAQVNTIQ 60
Db 1 MKKRILSAVLVSGVTLSSTATLSAVKADDFDAQIASQDSKINNLTAAQQQAQAQVNTIQ 60
```

```
Qy 61 QVSALOTQQAELQAEQENORLEAQSATLGGQIQTLSSKIVARNESLKQARSAQKSNATSY 120
Db 61 QVSALOTQQAELQAEQENORLEAQSATLGGQIQTLSSKIVARNESLKQARSAQKSNATSY 120
Qy 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDQKAAVEQKQENQAAINTVAANQE 180
Db 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDQKAAVEQKQENQAAINTVAANQE 180
Qy 181 TIAQNTNALNTQQAQLEAAQNLQAEELTTAQDQKATLVAQKAAAEAAARQAAAAQAAEA 240
Db 181 TIAQNTNALNTQQAQLEAAQNLQAEELTTAQDQKATLVAQKAAAEAAARQAAAAQAAEA 240
Qy 241 KAAAEAKALOEQAQAQAANNNTQATDASDQQAADNTQAAQTGDSSTQSAQAQVNS 300
Db 241 KAAAEAKALOEQAQAQAANNNTQATDASDQQAADNTQAAQTGDSSTQSAQAQVNS 300
Qy 301 DQSTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360
Db 301 DQSTTATEAQPSSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNWNGGOWAA 360
Qy 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAYVTGVGGQIQVQEAANYAGNQSIGNYRGWF 420
Db 361 SAAAGYRVGSTPSAGAVAVWNDGGYGHVAYVTGVGGQIQVQEAANYAGNQSIGNYRGWF 420
Qy 421 NPGSVSYIYPN 431
Db 421 NPGSVSYIYPN 431
```

## RESULT 7

ADX37276  
ID ADX37276 standard; protein; 431 AA.

XX  
AC ADX37276;

XX  
DT 21-APR-2005 (first entry)

XX  
DE Streptococcus mutant glucan binding protein B variant #5.

XX  
KW immunogenicity; immune stimulation; Glucan binding protein-B;  
KW microparticle; major histocompatibility complex; tooth disease.

XX  
OS Streptococcus mutans.

XX  
PN US2005031633-A1.

XX  
PD 10-FEB-2005.

XX  
PF 09-MAR-2004; 2004US-00797821.

XX  
PR 13-APR-1998; 98US-0081550P.

PR 08-JAN-1999; 99US-0115142P.

PR 12-APR-1999; 99US-00290049.

PR 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

PR 07-MAR-2003; 2003US-00383930.

XX  
(SMIT/) SMITH D J.

PA (TAUB/) TAUBMAN M A.

XX  
PI Smith DJ, Taubman MA;

XX  
WPI; 2005-151644/16.

XX  
PT New composition comprising a fragment of a glucan binding protein-B

PT (GbpB) that binds to MHC class II protein, and a biocompatible

PT microparticle, useful for producing an antibody (claimed) for immunizing

PT mammals against dental caries.

XX  
PS Claim 3; SEQ ID NO 33; 73pp; English.

XX  
CC The invention relates to a composition comprising a fragment of a glucan





OS Streptococcus mutans.  
XX US2005031633-A1.  
XX 10-FEB-2005.  
XX  
XX 09-MAR-2004; 2004US-00797821.  
XX  
XX 13-APR-1998; 98US-0081550P.  
XX 08-JAN-1999; 99US-0115142P.  
XX 12-APR-1999; 99US-00290049.  
XX 07-MAR-2002; 2002US-0363209P.  
XX 08-AUG-2002; 2002US-0402483P.  
XX 07-MAR-2003; 2003US-00383930.  
XX  
XX (SMIT/) SMITH D J.  
XX (TAUB/) TAUBMAN M A.  
XX  
XX Smith DJ, Taubman MA;  
XX WPI; 2005-151644/16.  
XX  
XX New composition comprising a fragment of a glucan binding protein-B  
XX (GbpB) that binds to MHC class II protein, and a biocompatible  
XX microparticle, useful for producing an antibody (claimed) for immunizing  
XX mammals against dental caries.  
XX  
XX Claim 3; SEQ ID NO 31; 73pp; English.  
XX  
XX The invention relates to a composition comprising a fragment of a glucan  
XX binding protein-B (GbpB) and a biocompatible microparticle, where the  
XX fragment binds to a major histocompatibility complex (MHC) class II  
XX protein. The composition is useful for producing an antibody for  
XX immunizing mammals against dental caries. This sequence corresponds to a  
XX Streptococcus mutans GbpB protein of the invention.  
XX  
XX Sequence 432 AA;  
XX  
Query Match 98.8%; Score 2079.5; DB 9; Length 432;  
Best Local Similarity 98.8%; Pred. No. 7.9e-133;  
Matches 427; Conservative 2; Mismatches 2; Indels 1; Gaps 1;  
Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60  
Db 1 MKKRILSAVLVSGVTLSATTLSAIAKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60  
Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQARSQAQKSNAAATSY 120  
Db 61 QVSALOTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQARSQAQKSNAAATSY 120  
Qy 121 INAIINSKVSDAINRVSAREVVSANEKMLQQEQDKAAVEQKQENQAINTVAANQE 180  
Db 121 INAIINSKVSDAINRVSAREVVSANEKMLQQEQDKAAVEQKQENQAINTVAANQE 180  
Qy 181 TIAQNTNALNTQQAELQAEALNLQAEITTAQDQKATLVAQKAAAEAEARQAAAAA 240  
Db 181 TIAQNTNALNTQQAELQAEALNLQAEITTAQDQKATLVAQKAAAEAEARQAAAAA 240  
Qy 241 KAAAEAKALQEAQAAQVAA-NNNTQATDASDQAAADNTQAAQTGDSSTEQSAQAVNN 299  
Db 241 KAAAEAKALQEAQAAQVAA-NNNTQATDASDQAAADNTQAAQTGDSSTEQSAQAVNN 300  
Qy 300 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNMGNGQWA 359  
Db 301 SDOESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNMGNGQWA 360  
Qy 360 ASAAAAGRYGVSTPSAGAVVWNDGGYGHVAYVTGVGGQIQVQEAANYAGNQSIGNYRGW 419  
Db 361 ASAAAAGRYGVSTPSAGAVVWNDGGYGHVAYVTGVGGQIQVQEAANYAGNQSIGNYRGW 420  
Qy 420 FNFQSVSYIYPN 431  
Db 421 FNFQSVSYIYPN 432

RESULT 10  
ADD93652  
ID ADD93652 standard; protein; 432 AA.  
XX  
XX AC ADD93652;  
XX  
XX 29-JAN-2004 (first entry)  
XX DT  
XX DE Streptococcus mutans glucan binding protein-B.  
XX KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
XX OS Streptococcus mutans.  
XX PN WO2003075845-A2.  
XX  
XX 18-SEP-2003.  
XX  
XX 07-MAR-2003; 2003WO-US006962.  
XX PF  
XX 07-MAR-2002; 2002US-0363209P.  
XX PR 08-AUG-2002; 2002US-0402483P.  
XX (FORS-) FORSYTH INST.  
XX PA  
XX Smith DJ, Taubman MA;  
XX WPI; 2003-845091/78.  
XX GENBANK; AY046413.  
XX  
XX Composition useful as vaccines for dental caries comprises a fragment of  
XX a glucan binding protein-B binding to a major histocompatibility complex  
XX class II protein.  
XX  
XX Claim 5; Page 8; 49pp; English.  
XX  
XX The present sequence is the protein sequence of the glucan binding  
XX protein-B (GbpB) of Streptococcus mutans strain 3SN1. The sequence  
XX includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
XX invention provides immunogenic compositions and vaccines for dental  
XX caries. The compositions comprise major histocompatibility complex (MHC)  
XX class II protein-binding GbpB peptides, especially HLA-binding peptides,  
XX covalently linked with peptide subunits of a glucosyltransferase. The  
XX compositions are used in a claimed method of eliciting production of an  
XX antibody in a mammal. Diepitopic or multiepitopic polypeptides can be  
XX prepared synthetically or by recombinant DNA technology. Antibodies  
XX raised against MHC class II binding fragments of GbpB can be used in  
XX passive immunisation.  
XX  
XX Sequence 432 AA;  
XX  
Query Match 98.4%; Score 2070.5; DB 7; Length 432;  
Best Local Similarity 98.6%; Pred. No. 3.2e-132;  
Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;  
Qy 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60  
Db 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQG 60  
Qy 61 QVSALOTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQARSQAQKSNAAATSY 120  
Db 61 QVSALOTQQAELQAEORLEAQSATLGGQIOTLSSKIVARNESLKQARSQAQKSNAAATSY 120  
Qy 121 INAIINSKVSDAINRVSAREVVSANEKMLQQEQDKAAVEQKQENQAINTVAANQE 180  
Db 121 INAIINSKVSDAINRVSAREVVSANEKMLHQEQDKAAVEQKQENQAINTVAANQE 180  
Qy 181 TIAQNTNALNTQQAELQAEALNLQAEITTAQDQKATLVAQKAAAEAEARQAAAAA 240  
Db 181 TIAQNTNALNTQQAELQAEALNLQAEITTAQDQKATLVAQKAAAEAEARQAAAAA 240



CC The invention relates to a protein (ABP25413-ABP30895) from group B  
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS  
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in  
CC the specification. The proteins have antibacterial and antiinflammatory  
CC activity. (I), nucleic acids encoding (I), ABM66044-ABN71526 and  
CC antibodies that bind (I) are used in the manufacture of medicaments for  
CC the treatment or prevention of infection or disease caused by  
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.  
CC Nucleic acids encoding (I) are used to detect Streptococcus in a  
CC biological sample. (I) is used to determine whether a compound binds to  
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be  
CC used as a vaccine or diagnostic composition. The disease caused by  
CC Streptococcus that is prevented or treated may be meningitis. Nucleic  
CC acid encoding (I) may be used to recombinantly produce (I) and may be  
CC used in gene therapy. Antibodies to (I) are used for affinity  
CC chromatography, immunoassays, and distinguishing/identifying  
CC Streptococcus proteins  
XX  
SQ Sequence 447 AA;

Query Match 58.5%; Score 1231; DB 5; Length 447;  
Best Local Similarity 59.9%; Pred. No. 3.3e-75;  
Matches 276; Conservative 43; Mismatches 98; Indels 44; Gaps 9;  
QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAIQASODSKINNLTAAQQAAQAVNTIQG 60  
DB 1 MKKRILSAVLVSGVTTLGTAAV--TVNADDFDSKIATDSVINTLSGQAAQAVNTAIKG 58  
QY 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNAATSY 120  
DB 59 QVGALESQQSELEAQAQLEAVSQQLGQEIQTLSNKIVARNESLUKQVRSQAQGN-LTNY 117  
QY 121 INAIINSKVSDAINRVSAIRVVSVANEKMLQOQEDKAAVEQKQENQAINTVAANOE 180  
DB 118 INTILNSKVSDAINRVSAIRVVSVANEKMLQOQEDKAAVEQKQENQAINTVAANKQ 177  
QY 181 TIAQNTNALMTQAQLEAEALQINLQAEITTAQDQKATLVAQKAAAEAAQAAQAAAE 240  
DB 178 AIENKKAALATQRAQLEAAQLSALQITTVQNEKASLIQAKAQAEEAAKAAEAQAAEA 237  
QY 241 KAAAEAKALQEAQAAQVANNNTQATDASDQAAADNTQAQTDGSTE--QSAQAQVN 298  
DB 238 KAQAEEAKAAQESVAKAQAAA-----QVESATPTETVQTQPTBIKPSNLTATS 286  
QY 299 NSQQUESTTATEA-----OPSASSASTAA-----VAANTSSANTYPAGQ 336  
DB 287 SATTVATTATATNEPKVTPSVVTKAVEAPKAVVSTPRAVSKPVVRSYDSSNTYPMGQ 346  
QY 337 CTWGVKSLAPWGVNGYNGQGWAAASAAAGYRVGSTPSAGAVW--NDGGYGHVAVTG 394  
DB 347 CTWGAKSMAWGVNGYNGWNAQWGAASARAAGYSGVTTTPRVGAVAVWPYDGGYGHVAVVTS 406  
QY 395 V-OGGQIOVEANYAGNOSIGNYRGWFPN---GSVSVIYEN 431  
DB 407 VANNSSIQVMESNYAGNMSGNIGYRGWFPNPSASGSVYIYEN 447

RESULT 13  
ADU69524  
ID ADU69524 standard; protein; 447 AA.  
XX  
AC ADU69524;  
XX  
DT 10-FEB-2005 (first entry)  
XX  
DE S agalactiae hyperimmune serum reactive antigen seqid 219.  
XX  
KW immune stimulation; antigen; bacterial surface display;  
KW hyperimmune serum reactive antigen; vaccine; bacterial infection;  
KW antibacterial; infection.  
XX  
OS Streptococcus agalactiae.  
XX

PN WO2004099242-2-A2.  
XX 18-NOV-2004.  
XX  
XX 06-MAY-2004; 2004WO-EP004856.  
XX  
XX 07-MAY-2003; 2003EP-00450112.  
PR 28-NOV-2003; 2003EP-00450266.  
XX (INTE-) INTERCELL AG.  
XX  
PI Meinke A, Nagy E, Hanner M, Horky M, Kallenda S, Prustomersky S;  
XX WPI; 2004-821662/81.  
DR N-PSDB; ADU69307.  
XX  
PT New nucleic acid molecule encoding a hyperimmune serum reactive antigen,  
PT useful for the manufacture of a vaccine against Streptococcus agalactiae  
PT infection.  
XX

Claim 14; SEQ ID NO 219; 221pp; English.  
PS  
XX  
CC The invention describes an isolated nucleic acid molecule encoding a  
CC hyperimmune serum reactive antigen or its fragment. Also described are: a  
CC vector comprising the nucleic acid molecule; a host cell comprising the  
CC vector; a hyperimmune serum-reactive antigen comprising a sequence  
CC encoded by the nucleic acid molecule and consisting of e.g., 85, 299, 467  
CC or 812 amino acids; fragments of hyperimmune serum-reactive antigens  
CC consisting of peptides comprising e.g., 76, 134, 221 or 576 amino acids;  
CC a process for producing a Streptococcus agalactiae hyperimmune serum  
CC reactive antigen; a process for producing a cell that expresses a S.  
CC agalactiae hyperimmune serum reactive antigen; a pharmaceutical  
CC composition, especially a vaccine, comprising the hyperimmune serum-  
CC reactive antigen or nucleic acid molecule; an antibody that binds at  
CC least to a selective part of the hyperimmune serum-reactive antigen; a  
CC hybridoma cell line, which produces the antibody; a method for producing  
CC the antibody; an antagonist that binds to the hyperimmune serum-reactive  
CC antigen; a method for identifying an antagonist capable of binding to the  
CC hyperimmune serum-reactive antigen; a method for identifying an  
CC antagonist capable of reducing or inhibiting the interaction activity of  
CC a hyperimmune serum-reactive antigen to its interaction partner; a  
CC process for in vitro diagnosing a disease related to expression of the  
CC hyperimmune serum-reactive antigen; and a process for in vitro diagnosis  
CC of a bacterial infection, especially a S. agalactiae infection. The  
CC hyperimmune serum reactive antigen is useful for isolating, purifying  
CC and/or identifying an interaction partner of the hyperimmune serum  
CC reactive antigen. The hyperimmune serum reactive antigen is useful for  
CC generating a peptide comprising anticalines, or for the manufacture of a  
CC functional nucleic acid comprising aptamers or Spiegelmers. The nucleic  
CC acid molecule is useful for the manufacture of a functional ribonucleic  
CC acid comprising ribozymes, antisense nucleic acids or siRNA. The nucleic  
CC acid molecule, hyperimmune serum-reactive antigen or antibody is useful  
CC for the manufacture of a vaccine against S. agalactiae infection. This is  
CC the amino acid sequence of a Streptococcus agalactiae hyperimmune serum  
CC reactive antigen.  
XX  
SQ Sequence 447 AA;

Query Match 58.5%; Score 1231; DB 8; Length 447;  
Best Local Similarity 59.9%; Pred. No. 3.3e-75;  
Matches 276; Conservative 43; Mismatches 98; Indels 44; Gaps 9;  
QY 1 MKKRILSAVLVSGVTLSSTATTLSAVKADDFDAIQASODSKINNLTAAQQAAQAVNTIQG 60  
DB 1 MKKRILSAVLVSGVTTLGTAAV--TVNADDFDSKIATDSVINTLSGQAAQAVNTAIKG 58  
QY 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLSSKIVARNESLKQQAARSQAQNAATSY 120  
DB 59 QVGALESQQSELEAQAQLEAVSQQLGQEIQTLSNKIVARNESLUKQVRSQAQGN-LTNY 117  
QY 121 INAIINSKVSDAINRVSAIRVVSVANEKMLQOQEDKAAVEQKQENQAINTVAANOE 180  
DB 118 INTILNSKVSDAINRVSAIRVVSVANEKMLQOQEDKAAVEQKQENQAINTVAANKQ 177

Db 118 INTILSKSVSDAVNRVVAIREVVSANEKMLAQOEADKALEAKQIENQNAINTVAANKQ 177

Qy 181 TIAQNTNALNTQQAQLEAAQINLQAEILTTAQDOKATILVAOKAAAEAAEAQAAAAQAAEA 240

Db 178 AIENKKAALATQRAQLEAAQLELSAQUTTVQNEKASLIQAKAAEAARAAEAQAAEA 237

Qy 241 KAAAEAKALQEQAAQAAQVAANNNTQATDASDQQAADNTQAAQTGDSTE--OSAAQAVN 298

Db 238 KAQAEAKAQAESVAKAQAA-----QVESATAPTETVQTQPRTEIKPSNLTATS 286

Qy 299 NSDOESTTATEA-----QPSASSASTAA-----VAANTSSANTYPAGQ 336

Db 287 SATTVAITTTATATNEPKVTQPSVVTKAVEAPKAVVSTPRAVSKPVVRSYDSSNTYPMGQ 346

Qy 337 CTWGVKSLAPWGVNGYNGGOWASAAAGYRVGSTPSAGAVVW--NDGGYGHVAVVTG 394

Db 347 CTWGAKSMAWGVNGYNGWGNANQWASARAAGYSVGTTPRVGAVAVWPYDGGGYGHVAVVTS 406

Qy 395 V-QGGQIQVQEQEANYAGNOSIGNYRGWFPN---GSVSYYIPN 431

Db 407 VANNSSIQWESYAGNWSIGNYRGSFNPSAGSVYYIYPN 447

RESULT 14

ADV88392

ID ADV88392 standard; protein; 447 AA.

AC ADV88392;

XX 24-FEB-2005 (first entry)

XX Streptococcus agalactiae protein sequence, SEQ ID 786.

XX Antibacterial; Vaccine; bacterial infection.

XX Streptococcus agalactiae.

XX FR2824074-A1.

XX 31-OCT-2002.

XX 26-APR-2001; 2001PR-00005642.

XX 26-APR-2001; 2001PR-00005642.

XX (INSP ) INST PASTEUR.

XX (CNRS ) CNRS CENT NAT RECH SCI.

XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;

XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu CP, Kunst F;

XX WPI; 2004-101891/11.

XX Genomic nucleotide sequences encoding polypeptides of Streptococcus

XX agalactiae for the development of vaccines, diagnostic tools, DNA chips

XX and identification of therapeutic targets.

XX Claim 6; SEQ ID NO 786; 2687pp; French.

XX The present invention relates to novel Streptococcus agalactiae

XX nucleotide sequences (I; ADV87607-ADV87745) and novel polypeptides (II;

XX CC ADV87746-ADV89950). The nucleotide sequences encode polypeptides of S.

XX agalactiae involved in the synthesis of amino acids, cell membranes,

XX intermediate (central) metabolism, energetic metabolism, fatty acid and

XX phospholipid metabolism, nucleotide metabolism including purines,

XX pyrimidines and/or nucleosides, regulatory functions, replication,

XX transcription, translation, protein transport, adaptation to atypical

XX conditions, sensitivity to medicines and/or analogues, functions related

XX to transposons, biosynthesis of cofactors, prosthetic groups and

XX transporters, cell membrane proteins and cellular machinery. (I) are

XX useful for the detection and/or amplification of nucleic acids.

XX Pharmaceutical composition comprising (I) or (II) are useful for

XX treatment of a bacterial S. agalactiae infection. Note: WO200292818A2 is

CC equivalent for the present basic patent FR2824074A1. WO200292818A2

CC contains 6617 sequence whereas the present patent only contains 2344

CC sequences.

XX Sequence 447 AA;

Qy 1 MKKRILSVLVSVSVTSSATTLTSAVKADPDPAQIASODSKINNLTQAOQAAQAVNTIQG 60

Db 1 MKKRILSVLVSVSVTTLGTTAAV--TVNADDFDSKIAATSDSVINTLSGQQAQAQVTAIKG 58

Qy 61 QVSALOTQQAELCAENORLEAQSATLGGQIQTLSSKIVARNESILKQOARSAQKSNAAATSY 120

Db 59 QVGALSSQSELEAQAQNAQLEAVSQLGQELQTLSSKIVARNESILKQOARSAQKGN-ITNY 117

Qy 121 INAINSKSVSDAINRVSAIREVVSANEKMLQOEADKALEAKQIENQNAINTVAANKQ 177

Db 118 INTILSKSVSDAVNRVVAIREVVSANEKMLAQOEADKALEAKQIENQNAINTVAANKQ 177

Qy 181 TIAQNTNALNTQQAQLEAAQINLQAEILTTAQDOKATILVAOKAAAEAAEAQAAAAQAAEA 240

Db 178 AIENKKAALATQRAQLEAAQLELSAQUTTVQNEKASLIQAKAAEAARAAEAQAAEA 237

Qy 241 KAAAEAKALQEQAAQAAQVAANNNTQATDASDQQAADNTQAAQTGDSTE--OSAAQAVN 298

Db 238 KAQAEAKAQAESVAKAQAA-----QVESATAPTETVQTQPRTEIKPSNLTATS 286

Qy 299 NSDOESTTATEA-----QPSASSASTAA-----VAANTSSANTYPAGQ 336

Db 287 SATTVAITTTATATNEPKVTQPSVVTKAVEAPKAVVSTPRAVSKPVVRSYDSSNTYPMGQ 346

Qy 337 CTWGVKSLAPWGVNGYNGGOWASAAAGYRVGSTPSAGAVVW--NDGGYGHVAVVTG 394

Db 347 CTWGAKSMAWGVNGYNGWGNANQWASARAAGYSVGTTPRVGAVAVWPYDGGGYGHVAVVTS 406

Qy 395 V-QGGQIQVQEQEANYAGNOSIGNYRGWFPN---GSVSYYIPN 431

Db 407 VANNSSIQWESYAGNWSIGNYRGSFNPSAGSVYYIYPN 447

RESULT 15

ADV81808

ID ADV81808 standard; protein; 447 AA.

XX ADV81808;

XX 24-FEB-2005 (first entry)

XX Streptococcus agalactiae protein, SEQ ID 2949.

XX Antibacterial; vaccine; bacterial infection.

XX Streptococcus agalactiae.

XX WO200292818-A2.

XX 21-NOV-2002.

XX 26-APR-2002; 2002WO-IB003059.

XX 26-APR-2001; 2001PR-00005642.

XX (INSP ) INST PASTEUR.

XX (CNRS ) CNRS CENT NAT RECH SCI.

XX Glaser P, Rusniok C, Chevalier F, Frangeul L, Lalioui L;

XX Zouine M, Couve E, Buchrieser C, Poyart C, Trieu-Cuot P, Kunst F;

XX WPI; 2004-101891/11.

XX



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

QM protein - protein search, using sw model

Run on: February 10, 2006, 22:20:08 ; Search time 50.8337 Seconds  
(without alignments)  
815.787 Million cell updates/sec

Title: US-10-797-821-29  
Perfect score: 2104  
Sequence: 1 MKKRILSAVLVSGVTLSSAT.....SIGNRGWFGNPGSVSYIYPN 431

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_80.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1057.5	50.3	392	2 G95258	secreted 45 kd pro
2	1057.5	50.3	392	2 B98124	general stress pro
3	672	31.9	461	2 JN0097	secreted 45K prote
4	658.5	31.3	456	2 B86903	hypothetical prote
5	544	25.9	507	2 S05542	hypothetical prote
6	359.5	17.1	398	2 AC1763	peptidoglycan lyti
7	346	16.4	401	2 A11387	cell wall-binding
8	317.5	15.1	473	2 F70031	conserved hypothet
9	264	12.5	581	2 E75383	cell wall binding
10	259	12.3	436	2 AH1387	cell wall-binding
11	257.5	12.2	437	2 AB1763	cell wall-binding
12	255.5	12.1	461	2 H84099	conserved hypothet
13	234.5	11.1	528	2 B75310	tolA protein - Esc
14	230.5	11.0	421	2 JV0057	probable tail fibe
15	227.5	10.8	971	2 B90835	probable membrane
16	227.5	10.8	973	2 C85693	membrane spanning
17	226	10.7	394	2 F90725	membrane spanning
18	226	10.7	394	2 G85576	hypothetical prote
19	225.5	10.7	166	2 C90029	hypothetical prote
20	222.5	10.6	267	2 F90028	hypothetical prote
21	222.5	10.6	392	2 F96937	cell wall-associat
22	222.5	10.6	492	2 A28616	M5 protein precurs
23	221.5	10.5	265	2 B89837	hypothetical prote
24	221.5	10.5	1122	2 G64887	probable tail fibe
25	219.5	10.4	4776	2 E95206	cell wall surface
26	218	10.4	255	2 G90061	hypothetical prote
27	214	10.2	1156	2 T34852	probable secreted
28	213.5	10.1	483	2 A26297	M6 protein - Strept
29	211	10.0	536	2 D84325	Hcr17 transducer l

RESULT 1  
G95258  
secreted 45 kd protein [imported] - Streptococcus pneumoniae (strain TIGR4)  
C/Species: Streptococcus pneumoniae  
C/Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C/Accession: G95258  
R/Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 295, 498-506, 2001  
A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A/Reference number: A95000; MUID:21357209; PMID:11463916  
A/Accession: G95258  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-392 <R>  
A/Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AE005672;  
A/Experimental source: strain TIGR4  
C/Genetics:  
A/Gene: SP2216

ALIGNMENTS

Query Match		50.3%	Score 1057.5;	DB 2;	Length 392;
Best Local Similarity		50.8%;	Pred. No. 2.1e-44;		
Matches 221;		Conservative 74;	Mismatches 89;	Indels 51;	Gaps 6;
Qy	1	MKKRILSAVLVSGVTLSSATLSAVKADDFDAQTASODSKINNLTACQQAQAQVNTIQG 60			
Db	1	MKKKILASLLSLTVMVSVQAVLFTTAHAETDDKIAAQDNKISNLTAQQQBAQRQVDIQE 60			
Qy	61	QVSALOTQQAELQAEENORLEAQSATLQQQIQTLSSKIVARNESLKQQAQAQKSNAAATSY 120			
Db	61	QVSAIQAEQSNLQAEENLQAEKLEGEITELSKNIVSRNQSLEKQARSQNTGAVTSY 120			
Qy	121	INAIINSKVSDAINRVSAIRVVSANEKMLQQQEQDKAAVEQKQEQENQAINTVAANQE 180			
Db	121	INTIVNSKSITEAISRVAAEMSEIVSANNKMLEQQKADKKAISEKQVANDDAINTVIANQQ 180			
Qy	181	TTAQTNTALNTQAQLENAQINLQAEITTAQDQKATLVQAQAAEAARQAAAAQAAAEA 240			
Db	181	KLADDAQAALTTKQAEKKAELSAAEKATAEGEKASLLEOKAAEAEEAARAAVAEAAAYKE 240			
Qy	241	KAAAEAKALQEAQAQVAANNNTQATDASDQAAAAADNTQAAQTGDSSTOASAAQAVNNS 300			
Db	241	KRASQ-----QQSVLASANTNLT-----AQQVAVSSAAA----- 270			
Qy	301	DQESTTATEAOPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWGVNNGGQMAA 360			
Db	271	-----PVRKVRPTYS-----TNASSYPIGECTGWGKTLAPWAGDVGWNGAQWAT 315			
Qy	361	SAAAGYRVGCVSTPSAGAVAVNDGGYCHVAVYVTVQGG-GOIYOQEAANYAGNQSGINRGW 419			

conserved hypothet  
surface antigen A  
M24 protein precu  
proteophosphoglyc  
IGA-specific metal  
R27-2 protein - Tr  
FntB protein limpo  
hypothetical prote  
Htr8 transducer li  
IGA-specific metal  
M protein precurs  
cell surface antig  
M protein precurs  
cytosolic repetiti  
regulatory protein  
130K surface exclu

Db 316 SAAAGFRTGTPQVGAIACWNDGGYGHVAVVTAVESTTTRIQVSESYNAGNRTIGNHRGW 375

Qy 420 FNP-----GSVSYIY 429

Db 376 FNPFTTSEGFVTIY 390

RESULT 2

B98124

general stress protein GSP-781 [imported] - Streptococcus pneumoniae (strain R6)

C;Species: Streptococcus pneumoniae

C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004

C;Accession: B98124

R;Hoskins, J.A.; Albion Jr., W.; Arnold, J.; Blaszczak, L.; Burgett, S.; DeHoff, B.S.; B

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

y, P.; Sun, P.M.; Winkler, M.E.

J. Bacteriol. 183, 5709-5717, 2001

A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.

A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A;Reference number: A97872; MUID:21429245; PMID:11544234

A;Accession: B98124

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-392 <KUR>

A;Cross-references: UNIPROT:Q97N55; UNIPROT:Q8DMY4; UNIPARC:UPI0000051B81; GB:AE007317;

C;Genetics:

A;Gene: gsp-781

Query Match 50.3%; Score 1057.5; DB 2; Length 392;

Best Local Similarity 50.8%; Pred. No. 2.1e-44;

Matches 221; Conservative 74; Mismatches 89; Indels 51; Gaps 6;

Qy 1 MKKRILSAVLVSGVTLLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

Db 1 MKKKILASLLSTVMSQVAVLTTAHAETDDKIAAQDNKISLNTAAQQAAQKQVDQIQ 60

Qy 61 QVSALOTQAAELOAENORLEAQATLGOQIQTLLSSKIVARNESLKKQARSQAQKSNATSY 120

Db 61 QVSAIQAEQSNLQANDRLQAESKLEGEITEISKNIVSRNSLEKQARSQAQTNGAVTSY 120

Qy 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEODKAAVEQKQENQAAINTVAANQE 180

Db 121 INTIVNSKSI TEAISRVAAEMSEIVSANNKMLEQKADKKAISEKVANNDAINVTIANQQ 180

Qy 181 TIAQNTNALNTQAAQLAEALNLQAEELTTAQDQKATLVAQKAAAEAEARQAAQAAEA 240

Db 181 KLADDAQAQLTTQAELEKAAELSLAAEKATAEGBKASLLEQKAAAEAEARAAVAEAAIKE 240

Qy 241 KAAAEAKALQEAQAAQVAANNNTQATDASDQAAAADNTQAAQTGDSTEQSAQAQVNN 300

Db 241 KRASQ-----QQSVLASANTNLT-----AQVQAVSESAAA----- 270

Qy 301 DQESTTATEAQPASASTAAVAANTSSANTYPAGQCTGWGKSLAPWGVNNGWGOWAA 360

Db 271 -----PVRKVRPTY-----TNASSYPIGECTGWGKTLAPWAGDVGNGAQMAT 315

Qy 361 SAAAAGVRVSTPSAGAVVWNDGGYGHVAVVTGVQG-GQIQVQEPANYAGNOSIGNYR 419

Db 316 SAAAAGFRTGTPQVGAIACWNDGGYGHVAVVTAVESTTTRIQVSESYNAGNRTIGNHRGW 375

Qy 420 FNP-----GSVSYIY 429

Db 376 FNPFTTSEGFVTIY 390

RESULT 3

JN0097

secreted 45K protein precursor - Lactococcus lactis

C;Species: Lactococcus lactis

C;Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999

C;Accession: JN0097

R;van Asseldonk, M.; Rutten, G.; Otteman, M.; Siezen, R.J.; de Vos, W.M.; Simons, G.

Gene 95, 155-160, 1990

A;Title: Cloning of usp45, a gene encoding a secreted protein from Lactococcus lactis su

A;Reference number: JN0097; MUID:91071599; PMID:2123812

A;Accession: JN0097

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-461 <VAN>

A;Cross-references: UNIPARC:UPI000016D72B; GB:M35374

Query Match 31.9%; Score 672; DB 2; Length 461;

Best Local Similarity 35.0%; Pred. No. 9.6e-26;

Matches 168; Conservative 104; Mismatches 140; Indels 68; Gaps 13;

Qy 1 MKKRILSAVLVSGVTLLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

Db 1 MKKKIISAILMSTVILSAAAPLSGVVADT-NSDIAKQDATISSAQSAKAQAQVDSLQS 59

Qy 61 QVSALOTQAAELOAENORLEAQATLGOQIQTLLSSKIVARNESLKKQARSQAQKSNATSY 120

Db 60 KVDLSQQKQTSTKQIAKIESEAKALNAQIATLINESIKERTKTLEAQRSQAQVNSSATNY 119

Qy 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEODKAAVEQKQENQAAINTVAANQE 180

Db 120 MDVAVNSKSLTDVIQKVATATVSSANKQLEQKEQKELSORSETVTVKYNVQFVLSQ 179

Qy 181 TIAQNTNALNTQAAQLAEALNLQAEELTTAQDQKATLVAQKAAAEAEARQAAQAAEA 240

Db 180 SLDSQAQELTSQAAELKVATLNVQATIAQDKKQALLDEKAAAEAKQAAYEA 239

Qy 241 KAAAEAKALQEAQAAQVAANNNTQATDASDQAAAADNTQAAQTGDSTEQ-----SAAQ 295

Db 240 Q-----QKEAAQQAASATAKAVEAATSSASASSQAPQVSTTDNTTTSNASAS 291

Qy 296 AVNNSDOESTATEAQPASASTAAVAANTSS-----ANTYP 333

Db 292 SSNSSSSSSSSSSSSSSSSSSSSSNAGNTNGTGTGTGTTGGSGINSPIGNPYA 351

Qy 334 AGQCT---WG-----VKSLAPWGVNNGWGOWAAASAAAAG--YRVGSTPSAGAV 380

Db 352 GGCTDYVMQYFAAQGYIRNIMP-----GNGGWASNGPAQGVLVHVGAAP--GVIAS 403

Qy 381 -----WNDGGYGHVAVVTGVQG-GQIQVQEPANYAGNOSIGNYRGNPWSVSIYPN 431

Db 404 SFSADFPVGYANSPYGHVAIVKSVNSDGTITIKEGGY-GTTWNGHER-TVSASGVTFLMPN 461

RESULT 4

E86903

hypothetical protein usp45 [imported] - Lactococcus lactis subsp. lactis (strain IL1403)

C;Species: Lactococcus lactis subsp. lactis

C;Date: 23-Mar-2001 #sequence\_revision 23-Mar-2001 #text\_change 09-Jul-2004

C;Accession: E86903

R;Bolotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Ehrlich

Genome Res. 11, 731-753, 2001

A;Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s

A;Reference number: A86625; MUID:21235186; PMID:11337471

A;Accession: E86903

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-456 <STO>

A;Cross-references: UNIPROT:Q9CDJ1; UNIPARC:UPI000000C6BFC; GB:AE005176; PID:gl2725296; P

A;Experimental source: strain IL1403

C;Genetics:

A;Gene: usp45

Query Match 31.3%; Score 658.5; DB 2; Length 456;

Best Local Similarity 35.1%; Pred. No. 4.3e-25;

Matches 165; Conservative 104; Mismatches 148; Indels 53; Gaps 14;

Qy 1 MKKRILSAVLVSGVTLLSSATTLISAVKADDFDAQIASQDSKINNLTAAQQAAQAQVNTIQ 60

Db 1 MKKKIISAILMSTVILSAAAPLSGVVADT-NSDIAKQDATISSAQSAKAQAQVDSLQS 59

Qy 61 QVSALOTQAAELOAENORLEAQATLGOQIQTLLSSKIVARNESLKKQARSQAQKSNATSY 120



[illegible]

```

Db      358  STNNNNNNNTVTPTPTPTTTPAPAPNPSGVSNGAAI VAEAYKYIGTPTV---WGGKD 414
Qy      356  -----GQWASAAAGAYRVG-STPSAGAVAVMND-GGYGHVAY 391
Db      415  PSGFDCSGFTRYVYLQVTRDIGWTVPOESAGTKISVQAKAGDLLFWGSGAGGTYHVAI 474
Qy      392  VTGVQGGQIQVQEAENVAGNOSIGNYRGWFNP 422
Db      475  SLG--GGQ-YIHAPQGENVKVGSVQ-WYTP 501

RESULT 6
AC1763
peptidoglycan lytic protein P45 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AC1763
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Blocker
.: Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.;
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krest, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;
ok, C.; Schluster, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1763
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <GLA>
A:Cross-references: UNIPROT:Q92748; UNIPARC:UPI000000CC935; GB:AL592022; PIDN:CAC97875.1;
A:Experimental source: strain Clip11262
A:Genetics:
A:Gene: spl

```

RESULT 7  
A11387  
peptidoglycan lytic protein P45 [imported] - *Listeria monocytogenes* (strain EGD-e)  
C:Species: *Listeria monocytogenes*  
C:Date: 27-Nov-2001 #sequence revision 27-Nov-2001 #text change 09-Jul-2004



```
Db      40 LAGTDTLSRRLEQLQRLIEQRRLSVQKQLFVRLRQLNAQQAARALDRDLATSE 99
      Qy      62 VSALQTQQAELQAEQNLEAQSATLGO---QIQTLSKIVARNESLKKQARSQAKS---N 115
      Db      100 VTDLLENLADVLA---RVTAATRALRETEAQIRVTRSQV---EALKVDARAVMKALYRA 152
      Qy      116 AATSYINAINSKSVSDAINRVSAIRVSVANERKMLQO-----QBQDKAAV 161
      Db      153 RNTQYMLLSQSNISDMLIRLDYANNMAGORNVEVMEQLRGAAAEALTQQLRQRQOSDAL 212
      Qy      162 EQKQENQ-----AAINTVAANOETIAQNTNALNTQQAQ----- 195
      Db      213 RGLQGEQTKLAEURDRRTQADALAEQRSACQQAQAVAVRTQQAALTQTTIDSLVGNV 272
      Qy      196 -----LEAAQLNLQAE-----LTTAQDQKATLVA-- 219
      Db      273 VRERTRLEERRRLEBEERRRREAEARRIREAQERARKEARLARIRAEQERKAREAEA 332
      Qy      220 --OKAAAEAEARQAAQAQAAAEAKAAAEAKALQE-----QAAQAVANNNTQ----- 265
      Db      333 ARQKALAE--ARAAQAKVAEAKARAEAKARAEQAQAQAQAQAQAQAQAQAQAQAQAQA 390
      Qy      266 --ATDASDQAAADNTQAAQTGDSFQSAQAQVN-----NSDQESTTATE 309
      Db      391 AAARAQOQQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA 450
      Qy      310 AQPSSASASTAAVAANTSSANTYPAGQCTWGVKSLAPVGVNGVGGQGWAAASAAAAGYRV 369
      Db      451 KQLAAELAPLPAISSRDLG---FPLP---GGKVAAP---YGTSGAQWEVLSCSGGRA 499
      Qy      370 GSTPSAGAVA-----VWMDGGVGHVAYVTGVGGQIQV 402
      Db      500 VAAQTGNVIAAYAAALGWILLDHGSNVITGVFGLQDTLLEV 542

RESULT 10
cell wall binding proteins homolog lmo2504 [imported] - Listeria monocytogenes (strain H
C;Species: Listeria monocytogenes
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AH1387
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AB1763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-437 <GLA>
A;Cross-references: UNIPROT:Q8V4E2; UNIPARC:UPI0000055221; GB:NC_003210; PIDN:CAD00582.1
A;Experimental source: strain EGD-e
C;Genetics:
A;Gene: lmo2504

Query Match      12.3%; Score 259; DB 2; Length 436;
Best Local Similarity 23.3%; Pred. No. 7.6e-06;
Matches 99; Conservative 86; Mismatches 160; Indels 76; Gaps 14;

Qy      6 LSAVLVSG-VTLSSATTLNAV-----KADFDQAIAQSODSKINNLTAAQQAQAV 55
Db      11 LSLIILIS-----APLTVKASINDMKRQSEIEQKSELKNLDTKSELNHLSENAE 70
Qy      56 NTIQGV-----SALQTQQAELQAEQNLEAQSATLGOQIQTLSKIVARNESLKKQARS 111
Db      71 ESLMKSLLDTNKKLKEQEDKVSSENEKLG---KLQKEMEKLRNDRQKVLNRAAI 126
Qy      112 QKSNATSYINALINSKSVSDAINRVSAIRVSVANERKMLQOQ---EQDKAAVQKQENQ 169
```

```
Db      127 QTTGTATSYLDMIFEAQDFKELVDVRTVTVVSAIVKADQINIMQDQDDQDKLVKVAESTSEKK 186
      Qy      170 AA-INTVAANOETIAQNTNALNTQQAELQAEQNLEAQSATLGOQKATLVAKAAAEBA 228
      Db      187 LENLKVLALELVSKNMESQKQEKNDLVMLAN-KKDLT--KSEQTLASEQCALTDDE 243
      Qy      229 RQAAAAQAAAEAKAAAEKALQEQAAQAQVAANNNTQATDASDQAAAAADNTQAAQTGDS 288
      Db      244 KRLASNTAGEKAKQEAIAKAAAEKRM-----QEAAAASSAKSA----- 281
      Qy      289 TEQSAAQAVNNSDQESTTATEAOPSASSASTAAVAANTSSANTYPVPA-GCTWGVKSLA-P 346
      Db      282 -----AVVKQPSSSSNEATEIVSSGGGQFIKPSGILTSFSESRNP 323
      Qy      347 WVGNYWGN-----GGQWAAASAAAAGYRVGVSTPSAGAVAVWMDGGVGHVAYVTGVGGQ 399
      Db      324 VTGKYESHKGQDIAGGCTVTVSAAASGTVVPSFGSGSGF---GGYGVVVKIDHNGFG 380

RESULT 11
AB1763
cell wall binding protein homolog lin2647 [imported] - Listeria innocua (strain Clip1126
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AB1763
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duseurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AB1763
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-437 <GLA>
A;Cross-references: UNIPROT:Q927Y9; UNIPARC:UPI000000CC934; GB:ALU592022; PIDN:CAC97874.1;
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin2647

Query Match      12.2%; Score 257.5; DB 2; Length 437;
Best Local Similarity 24.2%; Pred. No. 9e-06;
Matches 106; Conservative 84; Mismatches 137; Indels 111; Gaps 18;

Qy      6 LSAVLVSGVTLSATTLNAVKAQD-----DFDAQIAQSODSKINNLTAAQ 47
Db      11 LSLIILIS-----APLTVKASINDMKRQSEIEQKSELKNLDTKSELNHLSENAE 63
Qy      48 QQAA---QAQVNTIQGVSAALQTQQAELQAEQNLEAQSATLGOQIQTLSKIVARNESL 104
Db      64 KDAAKELESLLNSIDETNKKLKEQEDKVSSENEKLG---KLKKEIEKLRNDRQKVL 119
Qy      105 KQARSQAQKNAATSYINALINSKSVSDAINRVSAIRVSVANERKMLQOQEQDKAAVE-- 162
Db      120 DSRARAIQTGTATSYLDIMFEADDFKELIDRVTVVSAIVKADQINIMQDQDKDKLQVA 179
Qy      163 ----OKQOENQAAINTVAANOETIAQNTNALNTQQAELQAEQNLEAQSATLGOQKATLV 218
Db      180 ENSEKKLEN---LKVLALELVSRNNMESQKQEKNDLVMLAN-KKDLT--KSEQTLT 233
Qy      219 AQKAAAEAEARQAAQAQAAAEAKAAAEKALQEQAAQAQVAANNNTQATDASDQAAAAAD 278
Db      234 NEQGALSDEQKKLASNIAGEKAKQEAIAKAAAEKRM-----QEAAAAKN 276
Qy      279 NTQAAQTGDSFQSAQAQVNNSDQESTTATEAOPSA--SSASTAAVAANTSSANTY--PA 334
Db      277 ATVAAQ-----QPSSVTSAGTGATDVTSSGGGQFIKPA 310
Qy      335 -GCTWGVKSLA-PWVGNVWGN-----GGQWAAASAAAAGYRV-----GSTPSAGAVAV 381
Db      311 SGMLTSGFSERTNPVTGKYESHKGQDIAGGCTVTVSAAASGRVTVFSGFGATGS----- 363
```



F;14-34/Domain: transmembrane #status predicted <MSS>  
F;78-301/Domain: helical #status predicted <HSR>  
F;355-362/Region: nucleotide-binding motif A (P-loop)

Query Match 11.0%; Score 230.5; DB 2; Length 421;  
Best Local Similarity 28.2%; Pred. No. 0.00017;  
Matches 116; Conservative 67; Mismatches 177; Indels 51; Gaps 19;

QY 5 ILSAVLVGVTLSATTLSAVKADDDFAQIASQ-----DSKINNLTAAQQAQAQVNTIQ 59  
DB 16 IISAVL--HVILFAALIWSS-----FDENIEASAGGGSSSIDAVMVDSGAVVEQYKRMQ 68  
QY 60 GQVSA-----LQTOQA--ELQ-----AENORL-BAQSATLGGQIQTLSKIVARNES 103  
DB 69 SOESSAKRSDEQRMKKEQQAEELEKQAAEQERLAKOLEKRLAAQOBKQQAEEAAKQAE 128  
QY 104 LKO-QARSAQKSNAAATSYINAIINSKVSVDAINRVSAIREVVSANERKWLQQQODKAAVE 162  
DB 129 LKQQAEEAAKAAADAKAAEADAKAAEEAAK-----AAADAKKAAEAAKAAAE 181  
QY 163 -KQEQENQAAI--NTVAANOETIAQNTNALNTQQAQLEAAQLNLQAL-TTAODQKATLV 218  
DB 182 AOKKAEAAALKKKAEAAEAAAEAKKATEAEAKAEKAAEAKAAADKKA--A 239  
QY 219 AOKAAAE-EAARQAAAQAAAEAKAAEAKALQEQAAQAQVANNNT---QATDASDQQA 274  
DB 240 AERAAADKKAEEAAEAKAAADKAAEAKAAADKAAEAKAAEAKAAEAAADDFG 299  
QY 275 AADNTQAQOTGSTEQSAQAQVNNSDQESTTATEAQ-PSASSASTAAVAANTSSANTYP 333  
DB 300 ELSGKKNAPTGGAGKNNASPGSGNTKNGASGADINNYAGQIKSAIESKPYDASSYA 359  
QY 334 AGCTGWKSLAP--WVGNYWNGGOWA-ASAAAAGYRVGTPSAGAVAVW 381  
DB 360 GKCTCLRIK-LAPDGMLLDKPEGDPALCQALAAAKLAKPKPPQAVY 409

RESULT 15

B90835  
probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, substrain RIM  
C;Species: Escherichia coli  
C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C;Accession: B90835  
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen  
A;Reference number: A99629; MUID:21156231; PMID:11258796  
A;Accession: B90835  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-971 <HAY>  
A;Cross-references: UNIPROT:Q8XD04; UNIPARC:UPI000016542B; GB:BA0000007; PIDN:BA035073.1;  
A;Experimental source: strain O157:H7, substrain RIMD 0509952  
C;Genetics:  
A;Gene: ECs1650

Query Match 10.8%; Score 227.5; DB 2; Length 971;  
Best Local Similarity 25.5%; Pred. No. 0.00058;  
Matches 92; Conservative 58; Mismatches 172; Indels 39; Gaps 8;

QY 19 ATTLSAVKADDDFAQIASQDSKINNLTAAQQAQAQVNTIQGVVSALQTQQAELQENQR 78  
DB 121 AQNTAAKKSASDASTSAREATHATDAASARAATSAGQAASSAQSSASAGTASTKA 180  
QY 79 LEAQSATLGGQIQTLSKIVARNESLKQARSAQKSNAAATSYINAIINSKVSVDAINRV 138  
DB 181 TEA-----SKSAAAESSKSAAT--AKTSETNAAVSQSAA-----TS 221  
QY 139 AIREVVSANEKMLQQEQEODKAAVEKQEQENQAINTVAANQETIAQNTNALNTQQAQLEA 198  
DB 222 ASTATTKASAEASASDASAKAEKSSSETSAASSASSA-----ASSATAAGNSAKAAKT 276

Search completed: February 10, 2006, 22:33:21  
Job time : 53.9587 secs

QY 199 AQLNLQAEELTTAODQKATLVAKAAAEAEARQAAAAQAAAAAKAAAKALQEQAAQAV 258  
DB 277 SETNAKSSETAARQCSASAAAAAGSKTAAALSAASASTSAGQASATAGKGAESAASAST 336  
QY 259 AANNTOATDASDQQAADNDTQAAOTGDSSTEQSAAQAVNNSDQESTTTATEAOPSSASS 318  
DB 337 A---TTKAGEATQGAASASASAAKKTSETN---AKASETSAESSKTA-----AASSAS 384  
QY 319 TAAVAANTSSANTYPAGQCTWGVKSLAPWYNGYNGGOWAASAAAAAGYRVGTPSAGAV 378  
DB 385 SAASSASSASASKDEATROASAAKSSATTAST---KATEAAGSATATAAQAQSKSTAESAATR 441  
QY 379 A 379  
DB 442 A 442

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:07:53 ; Search time 295.878 Seconds  
(without alignments)  
1027.730 Million cell updates/sec

Title: US-10-797-821-29  
Perfect score: 2104  
Sequence: 1 MKKRILSAVLVSGVTLSAT.....SIGNRGWFNPGSVSYIYPN 431

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2104	100.0	431	2	Q38V3_STRMU
2	2100	99.8	431	2	Q9AG98_STRMU
3	2096	99.6	431	2	Q8DMW3_STRMU
4	2092	99.4	431	2	Q38V0_STRMU
5	2079.5	98.8	432	2	Q38V2_STRMU
6	2070.5	98.4	432	2	Q38V1_STRMU
7	1231	58.5	447	2	Q9AKA4_STRAG
8	1231	58.5	447	2	Q8E2H1_STRAS
9	1231	58.5	447	2	Q8E7X9_STRAS
10	1098.5	52.2	485	2	Q5M212_STRT1
11	1097	52.1	474	2	Q5M6K4_STRT2
12	1091.5	51.9	398	2	Q5XEL1_STRP6
13	1091.5	51.9	398	2	Q7CN07_STRP8
14	1091.5	51.9	398	2	Q9A128_STRP3
15	1085.5	51.6	398	2	Q9A128_STRP5
16	1057.5	50.3	392	2	Q8DMY4_STRR6
17	1057.5	50.3	392	2	Q97N55_STRPN
18	947	45.0	211	2	Q2ASZ7_STRMU
19	665	31.6	461	1	USP45_LACLC
20	658.5	31.3	456	2	Q9CDJ1_LACLA
21	591	28.1	524	2	Q9K2J9_ENTFC
22	576	27.4	516	1	P54_ENTFC
23	560	26.6	576	2	Q9KJ13_ENTHR
24	551.5	26.2	482	2	Q5MSM6_STRT2
25	548.5	26.1	470	2	Q9RQK3_9LACT
26	527.5	25.1	482	2	Q93LK4_ENTFA
27	522	24.8	461	2	Q56SA7_STRTR
28	511	24.3	449	2	Q93LK3_ENTFA
29	435.5	20.7	211	2	Q9DVU8_STRMU
30	416.5	19.8	226	2	Q93RG6_STRIT
31	415.5	19.7	544	2	Q840X3_STRMU

32	413.5	19.7	544	2	Q840V8_STRMU	Q840V8 streptococc
33	410.5	19.5	544	2	Q840W6_STRMU	Q840W6 streptococc
34	409.5	19.5	544	2	Q8DUR7_STRMU	Q8DUR7 streptococc
35	404.5	19.2	129	2	Q5M137_STRT1	Q5M137 streptococc
36	403	19.2	169	2	Q8E3F4_STRAS	Q8E3F4 streptococc
37	396	18.8	169	2	Q8DXT4_STRAS	Q8DXT4 streptococc
38	359.5	17.1	398	2	Q927Y8_LISIN	Q927Y8 listeria in
39	347	16.5	401	2	Q71WS3_LISMF	Q71WS3 listeria mo
40	346	16.4	401	2	Q9RE04_LISMO	Q9RE04 listeria mo
41	333.5	15.9	604	2	Q6M552_CORGL	Q6M552 corynebacte
42	331	15.7	430	2	Q81HV2_BACCG	Q81HV2 bacillus ce
43	327.5	15.6	600	2	Q8NOA0_CORGL	Q8NOA0 corynebacte
44	322	15.3	453	2	Q8G539_BIFLO	Q8G539 bifidobacte
45	317.5	15.1	427	2	Q73DG0_BACCI	Q73DG0 bacillus ce

ALIGNMENTS

RESULT 1  
Q938V3\_STRMU  
ID Q938V3\_STRMU PRELIMINARY; PRT; 431 AA.  
AC Q938V3;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DB Glucan-binding protein B.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=SQ32;  
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;  
RT "Cloning of the gbpB gene from Streptococcus mutans.";  
RL J. Dent. Res. 79:224-224(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=SQ32;  
RX MEDLINE=21481971; PubMed=11598068;  
RA DOI=10.1128/TAI.69.11.6931-6941.2001;  
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,  
RA Duncan M.J.;  
RT "Cloning of the Streptococcus mutans gene encoding glucan binding  
RT protein B and analysis of genetic diversity and protein production in  
RT clinical isolates.";  
RL Infect. Immun. 69:6931-6941(2001).  
DR EMBL; AY046410; AAK94500.1; -; Genomic\_DNA.  
DR InterPro; IPR007921; CHAP.  
DR InterPro; IPR009148; SilB.  
DR Pfam; PF05257; CHAP; 1.  
DR PRINTS; PR01852; SIBAPROTEIN.  
DR PROSITE; PS00911; CHAP; 1.  
SQ SEQUENCE 431 AA; 44620 MW; 464FE3B563FB7E51 CRC64;

Query Match		100.0%;	Score 2104;	DB 2;	Length 431;
Best Local Similarity		100.0%;	Pred. No. 5.2e-86;		
Matches 431;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MKKRILSAVLVSGVTLSAT	SSAT	SSAT	SSAT
Db	1	MKKRILSAVLVSGVTLSAT	SSAT	SSAT	SSAT
Qy	61	QVSAQTQAELOAENORLEA	QATL	QATL	QATL
Db	61	QVSAQTQAELOAENORLEA	QATL	QATL	QATL
Qy	121	INAIINSKVSDAINRVS	AI	AI	AI
Db	121	INAIINSKVSDAINRVS	AI	AI	AI
Qy	180	INAIINSKVSDAINRVS	AI	AI	AI
Db	180	INAIINSKVSDAINRVS	AI	AI	AI
Qy	240	TIANTNTALNTQQAQLE	AAQ	AAQ	AAQ





DR	InterPro; IPR007921; CHAP.	DR	Infect. Immun. 69:6931-6941 (2001).
DR	InterPro; IPR009148; Siba.	DR	EMBL; AY046414; AAK94504.1; -; Genomic_DNA.
DR	Pfam; PF05257; CHAP; 1.	DR	InterPro; IPR007921; CHAP.
DR	PRINTS; PR01852; SIBAPROTEIN.	DR	InterPro; IPR009148; Siba.
DR	PROSITE; PS09111; CHAP; 1.	DR	Pfam; PF05257; CHAP; 1.
KW	Complete proteome.	DR	PRINTS; PR01852; SIBAPROTEIN.
SQ	SEQUENCE 431 AA; 44620 MW; 2D1CA685248CCD3E CRC64;	DR	PROSITE; PS09111; CHAP; 1.
		SQ	SEQUENCE 431 AA; 44650 MW; 05D38D8B8BC4609F CRC64;
Query Match 99.6%; Score 2096; DB 2; Length 431;			
Best Local Similarity 99.5%; Pred. No. 1.2e-85;			
Matches 429; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
Qy	1 MKKRLSAVLVSGVTLSATLTSVAKDDPDAQIASQDSKINNLTAAQQAQAQVNTIQ 60	Qy	1 MKKRLSAVLVSGVTLSATLTSVAKDDPDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
Db	1 MKKRLSAVLVSGVTLSATLTSVAKDDPDAQIASQDSKINNLTAAQQAQAQVNTIQ 60	Db	1 MKKRLSAVLVSGVTLSATLTSVAKDDPDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
Qy	61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSNAATSY 120	Qy	61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSNAATSY 120
Db	61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSNAATSY 120	Db	61 QVSALQTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSNAATSY 120
Qy	121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAINTVAANQE 180	Qy	121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAINTVAANQE 180
Db	121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAINTVAANQE 180	Db	121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQDKAAVEQKQENQAINTVAANQE 180
Qy	181 TIAQNTVALNTQQAELQAEOLNLAELTTAQDQKATLVAQKAAAEAAARQAAAAA 240	Qy	181 TIAQNTVALNTQQAELQAEOLNLAELTTAQDQKATLVAQKAAAEAAARQAAAAA 240
Db	181 TIAQNTVALNTQQAELQAEOLNLAELTTAQDQKATLVAQKAAAEAAARQAAAAA 240	Db	181 TIAQNTVALNTQQAELQAEOLNLAELTTAQDQKATLVAQKAAAEAAARQAAAAA 240
Qy	241 KAAAEAKALQEAQAQVAANNNTQATDASDQAAAAADNTQAAGTGSTQSAQAQVNS 300	Qy	241 KAAAEAKALQEAQAQVAANNNTQATDASDQAAAAADNTQAAGTGSTQSAQAQVNS 300
Db	241 KAAAEAKALQEAQAQVAANNNTQATDASDQAAAAADNTQAAGTGSTQSAQAQVNS 300	Db	241 KAAAEAKALQEAQAQVAANNNTQATDASDQAAAAADNTQAAGTGSTQSAQAQVNS 300
Qy	301 DQESTTATEAQAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVWGVNGGQWAA 360	Qy	301 DQESTTATEAQAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVWGVNGGQWAA 360
Db	301 DQESTTATEAQAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVWGVNGGQWAA 360	Db	301 DQESTTATEAQAQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPVWGVNGGQWAA 360
Qy	361 SAAAAGRVGSTPSAGAVAVWMDGGYGHVAVYVTGQGGQIQVQEAANYAGNQSIGNYRGWF 420	Qy	361 SAAAAGRVGSTPSAGAVAVWMDGGYGHVAVYVTGQGGQIQVQEAANYAGNQSIGNYRGWF 420
Db	361 SAAAAGRVGSTPSAGAVAVWMDGGYGHVAVYVTGQGGQIQVQEAANYAGNQSIGNYRGWF 420	Db	361 SAAAAGRVGSTPSAGAVAVWMDGGYGHVAVYVTGQGGQIQVQEAANYAGNQSIGNYRGWF 420
Qy	421 NPGSVSYIYPN 431	Qy	421 NPGSVSYIYPN 431
Db	421 NPGSVSYIYPN 431	Db	421 NPGSVSYIYPN 431
RESULT 4			
Q938V0_STRMU			
ID	Q938V0_STRMU PRELIMINARY;	ID	Q938V2_STRMU PRELIMINARY;
AC	Q938V0;	AC	Q938V2;
DT	01-DEC-2001 (TrEMBLrel. 19, Created)	DT	01-DEC-2001 (TrEMBLrel. 19, Created)
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)	DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)	DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE	Glucan-binding protein B.	DE	Glucan-binding protein B.
OS	Streptococcus mutans.	OS	Streptococcus mutans.
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC	Streptococcus.	OC	Streptococcus.
OX	NCBI_TaxID=1309;	OX	NCBI_TaxID=1309;
RN	[1]	RN	[1]
RP	NUCLEOTIDE SEQUENCE.	RP	NUCLEOTIDE SEQUENCE.
RC	STRAIN=5SM3;	RC	STRAIN=15JP2;
RA	Jin S., Duncan M.J., Taubman M.A., Smith D.J.;	RA	Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT	"Cloning of the gbpB gene from Streptococcus mutans.";	RT	"Cloning of the gbpB gene from Streptococcus mutans.";
RL	J. Dent. Res. 79:224-224 (2000).	RL	J. Dent. Res. 79:224-224 (2000).
RL	[2]	RL	[2]
RP	NUCLEOTIDE SEQUENCE.	RP	NUCLEOTIDE SEQUENCE.
RC	STRAIN=5SM3;	RC	STRAIN=15JP2;
RX	MEDLINE=21481971; PubMed=11598068;	RX	MEDLINE=21481971; PubMed=11598068;
RX	DOI=10.1128/IAI.69.11.6931-6941.2001;	RX	DOI=10.1128/IAI.69.11.6931-6941.2001;
RA	Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.;	RA	Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.;
RA	Duncan M.J.;	RA	Duncan M.J.;
RT	"Cloning of the Streptococcus mutans gene encoding glucan binding	RT	"Cloning of the Streptococcus mutans gene encoding glucan binding
RT	protein B and analysis of genetic diversity and protein production in	RT	protein B and analysis of genetic diversity and protein production in
RT	clinical isolates.";	RT	clinical isolates.";

```
RT clinical isolates.";
RL Infect. Immun. 69:6931-6941 (2001).
DR ENBL; AY046412; AAK94502.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44648 MW; E769B2504AE50E9 CRC64;

Query Match 98.8%; Score 2079.5; DB 2; Length 432;
Best Local Similarity 98.8%; Pred. No. 6.4e-85;
Matches 427; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
DB 1 MKKRILSAVLVSGVTLSSTLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60

QY 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
DB 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQOEDKAAVEQKQENQAAINTVAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQOEDKAAVEQKQENQAAINTVAANQE 180

QY 181 TIAQNTNALNTQQAELQAEALNLQAEELTTAQQDKATILVAQKAAAEAEARQAAAAQAAEA 240
DB 181 TIAQNTNALNTQQAELQAEALNLQAEELTTAQQDKATILVAQKAAAEAEARQAAAAQAAEA 240

QY 241 KAAAEAKALQEQAQAQVAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299
DB 241 KAAAEAKALQEQAQAQVAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300

QY 300 SDQESTTATEAQSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNVWNGGQWA 359
DB 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNVWNGGQWA 360

QY 360 ASAAAAGYRGSTPSAGAVVNDGGYGHVAYVTGVGGQIQVEANYAGNQSIGNYRGW 419
DB 361 ASAAAAGYRGSTPSAGAVVNDGGYGHVAYVTGVGGQIQVEANYAGNQSIGNYRGW 420

QY 420 FNPQSVSYIYPN 431
DB 421 FNPQSVSYIYPN 432

RESULT 6
Q938V1_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224 (2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RX MEDLINE=21491971; PubMed=11598068;
RX DOI=10.1128/JAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
RA Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding

RT protein B and analysis of genetic diversity and protein production in
RL clinical isolates.";
RL Infect. Immun. 69:6931-6941 (2001).
DR ENBL; AY046413; AAK94503.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 432 AA; 44652 MW; 3F88ECB9A1F3B4F CRC64;

Query Match 98.4%; Score 2070.5; DB 2; Length 432;
Best Local Similarity 98.6%; Pred. No. 1.1e-84;
Matches 426; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 1 MKKRILSAVLVSGVTLSSTLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
DB 1 MKKRILSAVLVSGVTLSSTLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60

QY 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120
DB 61 QVSALOTQQAELQAEORLEAQSATLGGQIQTLLSSKIVARNESLKQQAARSQAQKSNAAATSY 120

QY 121 INAIINSKVSDAINRVSAIREVVSAANEKMLQQOEDKAAVEQKQENQAAINTVAANQE 180
DB 121 INAIINSKVSDAINRVSAIREVVSAANEKMLHQEODKAAVEQKQENQAAINTVAANQE 180

QY 181 TIAQNTNALNTQQAELQAEALNLQAEELTTAQQDKATILVAQKAAAEAEARQAAAAQAAEA 240
DB 181 TIAQNTNALNTQQAELQAEALNLQAEELTTAQQDKATILVAQKAAAEAEARQAAAAQAAEA 240

QY 241 KAAAEAKALQEQAQAQVAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 299
DB 241 KAAAEAKALQEQAQAQVAA-NNNTQATDASDQAAAAADNTQAAQTGDSSTEQSAAQAVNN 300

QY 300 SDQESTTATEAQSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNVWNGGQWA 359
DB 301 SDQESTTATAQPSASSASTAAVAANTSSANTYPAGQCTWGVKSLAPWVGNVWNGGQWA 360

QY 360 ASAAAAGYRGSTPSAGAVVNDGGYGHVAYVTGVGGQIQVEANYAGNQSIGNYRGW 419
DB 361 ASAAAAGYRGSTPSAGAVVNDGGYGHVAYVTGVGGQIQVEANYAGNQSIGNYRGW 420

QY 420 FNPQSVSYIYPN 431
DB 421 FNPQSVSYIYPN 432

RESULT 7
Q9AKA4_STRAG
ID Q9AKA4_STRAG PRELIMINARY; PRT; 447 AA.
AC Q9AKA4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PcsB protein precursor.
GN Name=PcsB;
OS Streptococcus agalactiae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1311;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=6313;
RX MEDLINE=21101799; PubMed=11157929;
RX DOI=10.1128/JB.183.4.1175-1183.2001;
RA Reinscheid D.J., Gottschalk B., Schubert A., Eikmanns B.J.,
RA Chhatwal G.S.;
RT "Identification and molecular analysis of PcsB, a protein required for
cell wall separation of group B streptococcus.";
RL J. Bacteriol. 183:1175-1183 (2001).
DR ENBL; AJ277292; CAC28144.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
```

```

DR InterPro: IPR009148; Siba.
DR Pfam: PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PSS0911; CHAP; 1.
KW SIGNAL.
FT CHAIN 1 25 Potential.
FT SIGNAL 26 447 PeSB protein.
FT CHAIN 26 447 PeSB protein.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 58.5%; Score 1231; DB 2; Length 447;
Best Local Similarity 59.9%; Pred. No. 3.1e-47;
Matches 276; Conservative 43; Mismatches 98; Indels 44; Gaps 9;

QY 1 MKKRILSAVLVSGVTLSATTLGAVKADDFDAQIASODSKINNLTAAQQAAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTTLGTAAV--TVNADDFDSKIATDSVINTLSGQAAAQNVTAIKG 58

QY 61 QVSALOTQQAELQAEORLEAQAATLGGQQTQTLSSKIVARNESLKQAARSQAQNSAATSY 120
DB 59 QVGALSSQSELEAQAQNAQLAEAVSQQLGQEIQTLSNKIVARNESLKKQVRSQAQGN-LTNY 117

QY 121 INAIINSKSVSDAINRVSAIREVVVSANEKMLQQOEQDKAAVEQKQEQENQAQINVAANO 180
DB 118 INTILSKSVSDAVNRVVAIREVVVSANEKMLAQOEADKALEAKQIENQAINVAANKQ 177

QY 181 TIAQNTNALNTQQAELQAEALNLOAELTTAQQDKATLVAQKAAAEAAARQAAQAQAAEA 240
DB 178 AIENNKAAALATQRAQLAEALQELSAQLTTVQNEKASLIQAKAQAEEAARQAQAQAAEA 237

QY 241 KAAAEAKALQEQAAQAAQVAANNNTQATDASDQQAADNTQAAQTGDSSTE--QSAQAQVN 298
DB 238 KAQAEEAKAQAESVAKAQAQA-----QVESATAPTETVQTQPRTEIKPSNLTATS 286

QY 299 NSDOESTTATEA-----QPSASSASTAA-----VAANTSSANTYPAGQ 336
DB 287 SATTVATTATATNEPKVTQPSVVTKAVEAPKAVSSTPRAVSKPVVRSVDSSTNTPMGQ 346

QY 337 CTWGVKSLAPVWNGYNGWGQAAASAAAAGRYGVSTPSAGAVAW--NDGGYGHVAVVTG 394
DB 347 CTWGAKSMAWNGYNGWGNQWASARAAGYSVGTTPRVGAVAWPVDGGYGHVAVVTS 406

QY 395 V-OGGQIQVQEAANYAGNQSGNIGYRGWFPN---GSVSYIYPN 431
DB 407 VANNSSIQVMESYAGNMSGNYRGSGFNPSPASGSVYIYPN 447

RESULT 8
Q8E2H1_STRAS5 PRELIMINARY; PRT; 447 AA.
AC Q8E2H1;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PeSB protein.
GN Name=pscb; OrderedLocusNames=SAG0017;
OS Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216466;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547; DOI=10.1073/pnas.182380799;
RA Tettelin H., Maignani V., Cieslewicz M.J., Eisen J.A., Peterson S.N.,
RA Wessels M.R., Paulsen I.T., Nelson K.E., Margalit I., Read T.D.,
RA Madoff L.C., Wolf A.M., Beanan M.J., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radu D., Fedorova N.B., Scanlan D., Khouri H.M., Mulligan S.,
RA Carty H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,
RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Maione D.,
RA Rinaldo D., Rappuoli R., Telford J.B., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an

```

```

RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396(2002).
DR EMBL; AE014192; AAM98925.1; -, Genomic_DNA.
DR TIGR; SAG0017; -.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PRO1852; SIBAPROTEIN.
DR PROSITE; PSS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 447 AA; 46681 MW; F4DB14B0A5F962C8 CRC64;

Query Match 58.5%; Score 1231; DB 2; Length 447;
Best Local Similarity 59.9%; Pred. No. 3.1e-47;
Matches 276; Conservative 43; Mismatches 98; Indels 44; Gaps 9;

QY 1 MKKRILSAVLVSGVTLSATTLGAVKADDFDAQIASODSKINNLTAAQQAAQAQVNTIQG 60
DB 1 MKKRILSAVLVSGVTTLGTAAV--TVNADDFDSKIATDSVINTLSGQAAAQNVTAIKG 58

QY 61 QVSALOTQQAELQAEORLEAQAATLGGQQTQTLSSKIVARNESLKQAARSQAQNSAATSY 120
DB 59 QVGALSSQSELEAQAQNAQLAEAVSQQLGQEIQTLSNKIVARNESLKKQVRSQAQGN-LTNY 117

QY 121 INAIINSKSVSDAINRVSAIREVVVSANEKMLQQOEQDKAAVEQKQEQENQAQINVAANO 180
DB 118 INTILSKSVSDAVNRVVAIREVVVSANEKMLAQOEADKALEAKQIENQAINVAANKQ 177

QY 181 TIAQNTNALNTQQAELQAEALNLOAELTTAQQDKATLVAQKAAAEAAARQAAQAQAAEA 240
DB 178 AIENNKAAALATQRAQLAEALQELSAQLTTVQNEKASLIQAKAQAEEAARQAQAQAAEA 237

QY 241 KAAAEAKALQEQAAQAAQVAANNNTQATDASDQQAADNTQAAQTGDSSTE--QSAQAQVN 298
DB 238 KAQAEEAKAQAESVAKAQAQA-----QVESATAPTETVQTQPRTEIKPSNLTATS 286

QY 299 NSDOESTTATEA-----QPSASSASTAA-----VAANTSSANTYPAGQ 336
DB 287 SATTVATTATATNEPKVTQPSVVTKAVEAPKAVSSTPRAVSKPVVRSVDSSTNTPMGQ 346

QY 337 CTWGVKSLAPVWNGYNGWGQAAASAAAAGRYGVSTPSAGAVAW--NDGGYGHVAVVTG 394
DB 347 CTWGAKSMAWNGYNGWGNQWASARAAGYSVGTTPRVGAVAWPVDGGYGHVAVVTS 406

QY 395 V-OGGQIQVQEAANYAGNQSGNIGYRGWFPN---GSVSYIYPN 431
DB 407 VANNSSIQVMESYAGNMSGNYRGSGFNPSPASGSVYIYPN 447

RESULT 9
Q8E7X9_STRAS3 PRELIMINARY; PRT; 447 AA.
AC Q8E7X9;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE PeSB protein.
GN Name=pscb; OrderedLocusNames=gbs0016;
OS Streptococcus agalactiae (serotype III).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=216495;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NEM316 / Serotype III;
RX MEDLINE=22242508; PubMed=1254221;
RA Glaser P., Zouinok C., Buchrieser C., Chevalier F., Frangeul L.,
RA Meadek T., Rouine M., Couve E., Lailou L., Poyart C., Trieu-Cuot P.,
RA Kunst F.;
RT "Genome sequence of Streptococcus agalactiae, a pathogen causing
RT invasive neonatal disease.";
RL Mol. Microbiol. 45:1499-1513(2002).
DR EMBL; AL766843; CAD45661.1; -, Genomic_DNA.

```



```

RC STRAIN=LMG 18311;
RA Borges F., Lajey S., Thibessard A., Fernandez A., Gintz B., Hols P.,
RA Decaris B., Lebiond-Bourget N.;
RT "cse, a Chimeric and Variable Gene, Encodes an Extracellular Protein
RT Involved in Cellular Segregation in Streptococcus thermophilus.";
RL J. Bacteriol. 187:2737-2746(2005).
DR EMBL; CP000023; AAW59752.1; -; Genomic_DNA.
DR EMBL; AY730643; AAW82375.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 474 AA; 48142 MW; 1BACGA9E0A0A200D CRC64;

Query Match 52.1%; Score 1097; DB 2; Length 474;
Best Local Similarity 52.2%; Pred. No. 2.9e-41;
Matches 249; Conservative 67; Mismatches 111; Indels 50; Gaps 12;

Qy 1 MKKRILSAVLGVLTSSATLTSVAVKADDDFAQIASODSKINNLTAAQOAAQAQVNTIQG 60
Db 1 MKKRILSAVLGVLTSSAA---ASVHAEDYDSQIAATNAISNLASQOERAQAQVATIQS 57

Qy 61 QVSLQTOQAELQAEORLEAQSATLGOQIOTLSSKIVARNESLKQARSQAQKNAATSY 120
Db 58 QVSLTRQKTELEAKNAELEKVSADLESEIQELSSKIVARQDSLAKQARSQAQNNATSY 117

Qy 121 INAIINSKVSDAINRVSAIREVVVSAANEKMLQOQEQDKAAVEQKQENQAQAINVAAQOE 180
Db 118 INSILNSKSISEALTRITAKLSKVTVANNDLTKQESDQKELAAQOENQAQAINVAAQNS 177

Qy 181 TIAQNTNALNTQQAELAAQLNQAELTTAQDQKATLVAQKAAAEAAEQAAA---AQA 236
Db 178 ELETTEAGLTQQAELAAQVTLAELATQAQNEKTSLSVSAKSTAESVAASTAASVAQSOA 237

Qy 237 AAEKAAAEKALQEQEA---AQAOVAANNNTQA---TDASDQAAA---ADNTQAAQOTGD 287
Db 238 IAESEATQVVASSEAAATSVASSEVAATSEAVAQFSETPVSETSTASAEAAQEPASSETSE 297

Qy 288 STEGSAQAQVNN-----SDQESTTAT-----EAQPSASSASTAAVAANT 326
Db 298 VQPSAAPAVSEAPASVAPVATSEAPATSEAPASVAPVATSEAPASVAPVATSEAPATSEAP 357

Qy 327 SSA-----NTYPAGQCTGWGKSLAPWGVNGYNGGQWAAASAAAAGRYGVSTPSSAGAVAVW- 381
Db 358 VSAASTENTYPVGCQCTGWGKSLAPWAGNNGWNAKNIASQAACHSVGTTTPVAGAIAMP 417

Qy 382 ND-GGYGHVAVTGVQVQ-GQIQVOEANYAGNQSIGNYRGWPNP-----GSVSVIYP 430
Db 418 NDGGYGHVAVTVTSASGANSIQVMESYAGNMSISNYRGTFDPTSSAHGGSVFYIYP 474

RESULT 12
QXCELL STRP6 PRELIMINARY; PRT; 398 AA.
AC QXCELL;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Secreted protein.
GN OrderedLocusNames=M6_Spy0017;
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301450;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS10394;
RX PubMed=15272401; DOI=10.1086/422697;
RA Banks D.J., Porcella S.F., Barbican K.D., Beres S.B., Phillips L.E.,
RA Voyich J.M., DeLeo F.R., Martin J.M., Somerville G.A., Musser J.M.;
RT "Progress toward characterization of the group A Streptococcus

```

```

RT metagenome: complete genome sequence of a macrolide-resistant serotype
RT M6 strain.";
RL J. Infect. Dis. 190:727-738(2004).
DR EMBL; CP000003; AAT86152.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 42028 MW; 5DCEDA78CB863B60 CRC64;

Query Match 51.9%; Score 1091.5; DB 2; Length 398;
Best Local Similarity 53.9%; Pred. No. 4.3e-41;
Matches 233; Conservative 57; Mismatches 107; Indels 35; Gaps 6;

Qy 1 MKKRILSAVLGVLTSSATLTSVAVKADDDFAQIASODSKINNLTAAQOAAQAQVNTIQG 60
Db 1 MKKRILSAVLGVLTGAAATT---VGAEDLSTLAKQDSIISNLTTEQAQONQVSALQA 57

Qy 61 QVSLQTOQAELQAEORLEAQSATLGOQIOTLSSKIVARNESLKQARSQAQKNAATSY 120
Db 58 QVSSLOSEQDKLTARNTLEALSQRPEQEIKAFTSQIVARNEKLNQARSAYKNNETSGY 117

Qy 121 INAIINSKVSDAINRVSAIREVVVSAANEKMLQOQEQDKAAVEQKQENQAQAINVAAQOE 180
Db 118 INALLNSKISDVVNRVLVAINRAVSAANAKLEQOQKADKVSLEEKQAAANTQTAINTAANMA 177

Qy 181 TIAQNTNALNTQQAELAAQLNQAELTTAQDQKATLVAQKAAAEAAEQAAAQAQAAEA 240
Db 178 MAENQNTLTQQNLEAATANLALQASATEDKANLVQAEAAEKAAAEALAEQAAKV 237

Qy 241 KAAAEAKALQEQEAQAQVAVVNNNDGQVAVVTTGVQ-GGQIQVOEANYAGNQSIGNYRGW 300
Db 238 KA-----QEQA-----QQAASVEAAKSAITPAPQATPAAQSSNAI 273

Qy 301 DQESTTATEAQPSASSASTAAVAANTSANTYPAGQCTGWGKSLAPWGVNGYNGGQWAA 360
Db 274 EPAALTA---PAAPSARP---QTSYDSSNTYPVGCQCTGWGAKSLAPWAGNNGGQWAV 326

Qy 361 SAAAGYRVGVTSTPAGAVAVVNDGQVAVVTTGVQ-GGQIQVOEANYAGNQSIGNYRGW 419
Db 327 SAQAGYRTGSTPVGAIWVNDGGYGHVAVVVEVQSSASIRVWESYSGRYADHRGW 386

Qy 420 FNPGSVSIYPN 431
Db 387 FNPGTFTFIYP 398

RESULT 13
Q7CNO7 STRP8 PRELIMINARY; PRT; 398 AA.
AC Q7CNO7;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=spvM18_0020;
OC Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301451;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MGAS232 / Serotype M18;
RX MEDLINE=21927593; PubMed=11917108; DOI=10.1073/pnas.062526099;
RA Smoot J.C., Barbican K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Sylva G.L., Sturdevant D.B., Ricklefs S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Veasy L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks.";

```

```
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
DR EMBL; AE009684.1; -, Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009146; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS00911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 5DCEDA78CB863B60 CRC64;

Query Match 51.9%; Score 1091.5; DB 2; Length 398;
Best Local Similarity 53.9%; Pred. No. 4.3e-41;
Matches 233; Conservative 57; Mismatches 107; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
DB 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
QY 61 QVSALOTQAQLEAQAENORLEAQAQATLGGQIQTLSSKIVARNESLKQAQSAQKSNATSY 120
DB 61 QVSALOTQAQLEAQAENORLEAQAQATLGGQIQTLSSKIVARNESLKQAQSAQKSNATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEMKLOQEOODKAAVEKQOENQAINTVAANO 180
DB 121 INAIINSKVSDAINRVSAIREVVSANEMKLOQEOODKAAVEKQOENQAINTVAANO 180
QY 178 MABENQNTLTQQAQLEAQAENORLEAQAQATLGGQIQTLSSKIVARNESLKQAQSAQKSNATSY 237
DB 178 MABENQNTLTQQAQLEAQAENORLEAQAQATLGGQIQTLSSKIVARNESLKQAQSAQKSNATSY 237
QY 241 KAAAEAKALQEOQAQAQVAAANNNTQATDASDQQAADNTQAAQTGDSTEQSAQAQVNS 300
DB 241 KAAAEAKALQEOQAQAQVAAANNNTQATDASDQQAADNTQAAQTGDSTEQSAQAQVNS 300
QY 420 FNPQSVSYIYPN 431
DB 420 FNPQSVSYIYPN 431
QY 387 FNPQSVSYIYPN 398
DB 387 FNPQSVSYIYPN 398

RESULT 15
Q9A128_STRPY PRELIMINARY; PRT; 398 AA.
AC Q9A128; Q7BH59;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Secreted protein Siba precursor.
GN OrderedLocusNames=SPY0019;
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;

RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673 (2002).
DR EMBL; AE009684.1; -, Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009146; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS00911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 398 AA; 5DCEDA78CB863B60 CRC64;

Query Match 51.9%; Score 1091.5; DB 2; Length 398;
Best Local Similarity 53.9%; Pred. No. 4.3e-41;
Matches 233; Conservative 57; Mismatches 107; Indels 35; Gaps 6;

QY 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
DB 1 MKKRILSAVLVSGVTLSATLTSVAVKADDFDAQIASQDSKINNLTAAQQAQAQVNTIQ 60
QY 61 QVSALOTQAQLEAQAENORLEAQAQATLGGQIQTLSSKIVARNESLKQAQSAQKSNATSY 120
DB 61 QVSALOTQAQLEAQAENORLEAQAQATLGGQIQTLSSKIVARNESLKQAQSAQKSNATSY 120
QY 121 INAIINSKVSDAINRVSAIREVVSANEMKLOQEOODKAAVEKQOENQAINTVAANO 180
DB 121 INAIINSKVSDAINRVSAIREVVSANEMKLOQEOODKAAVEKQOENQAINTVAANO 180
QY 178 MABENQNTLTQQAQLEAQAENORLEAQAQATLGGQIQTLSSKIVARNESLKQAQSAQKSNATSY 237
DB 178 MABENQNTLTQQAQLEAQAENORLEAQAQATLGGQIQTLSSKIVARNESLKQAQSAQKSNATSY 237
QY 241 KAAAEAKALQEOQAQAQVAAANNNTQATDASDQQAADNTQAAQTGDSTEQSAQAQVNS 300
DB 241 KAAAEAKALQEOQAQAQVAAANNNTQATDASDQQAADNTQAAQTGDSTEQSAQAQVNS 300
QY 420 FNPQSVSYIYPN 431
DB 420 FNPQSVSYIYPN 431
QY 387 FNPQSVSYIYPN 398
DB 387 FNPQSVSYIYPN 398

RESULT 14
Q8P318_STRPY PRELIMINARY; PRT; 398 AA.
AC Q8P318; Q7CFL7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Putative secreted protein.
GN OrderedLocusNames=SPY0015, SPY0015, SPY0015;
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=301448;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SSI-1 / Serotype M3;
RX MEDLINE=22683278; PubMed=12799345; DOI=10.1101/gr.1096703;
RA Nakagawa I., Kurokawa K., Yamashita A., Nakata M., Tomiyasu Y.,
RA Okahashi N., Kawabata S., Yamazaki K., Shiba T., Yaeunaga T.,
RA Hayashi H., Hattori M., Hamada S.;
RT "Genome sequence of an M3 strain of Streptococcus pyogenes reveals a
RT large-scale genomic rearrangement in invasive strains and new insights
RT into phage evolution.";
RL Genome Res. 13:1042-1055 (2003).
RN [2]
RP NUCLEOTIDE SEQUENCE.
```

RX MEDLINE=21192684; PubMed=11296296; DOI=10.1073/pnas.071559398;  
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,  
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,  
RA Qian Y., Jia H.G., Najjar P.Z., Ren Q., Zhu H., Song L., White J.,  
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.E.;  
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RA Fagan P.K., Reinscheid D., Gottschalk B., Chhatwal G.S.;  
RT "Identification and characterization of a novel secreted protein from  
RT group A streptococcus.";  
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE006474; AAK33158.1; -; Genomic\_DNA.  
DR EMBL; AF319999; AAL73135.1; -; Genomic\_DNA.  
DR InterPro; IPR007921; CHAP.  
DR InterPro; IPR009148; Siba.  
DR Pfam; PF05257; CHAP; 1.  
DR PRINTS; PR01852; SIBAPROTEIN.  
DR PROSITE; PS50911; CHAP; 1.  
KW Complete proteome; Signal.  
FT SIGNAL 1 23  
SQ SEQUENCE 398 AA; 41899 MW; 28A9B3F7195E969B CRC64;  
  
Query Match 51.6%; Score 1085.5; DB 2; Length 398;  
Best Local Similarity 53.7%; Pred. No. 7.9e-41;  
Matches 232; Conservative 57; Mismatches 108; Indels 35; Gaps 6;  
  
QY 1 MKKRILSAVLVSGVTLSATTLSAVKADDFDAQIASODSKINNLTAAQQAQAQVNTIQG 60  
DB 1 MKKRILSAVLVSGVTILGAATT--VGAEDLSTKIADKDSIISNLTTEKAAQNVSAQ 57  
  
QY 61 QVSALQTQQAELQAEQNRLEAQSATLQQIQTLSKIVARNESLKQQAARSQAQKSAATSY 120  
DB 58 QVSSLSQEQDKLTARNTLEALSRRFEQEI KALTSQIVARNEKLQARSAYKNETSGY 117  
  
QY 121 INAIINSKVSDAINRVSAIREVVSANEKMLQQEQEDKAAVEQKQEQENQAINTVAANQ 180  
DB 118 INALLNSKSIDVVNRLVAINRAVSAKALLEQQAQKADKVSLEEKQAANQTAINTAANMA 177  
  
QY 181 TIAQNTVALNTQQAQLEAAQLNLQAEILTTAQDQKATLVAQKAAAEAAQAAAAQAAEA 240  
DB 178 MAENQNTLRTQQANILVAATANLALQASATEDKANLVAQKAAEAQAAEAQAAKV 237  
  
QY 241 KAAAEAKALQEQAAQAAQVAANNNTQATDASDQAAAAADNTQAAQTGDSQAAQAVNNS 300  
DB 238 KA-----QEQA-----QQAASVEAAKSAITPAPQATPAAQSSNAI 273  
  
QY 301 DQESTTATEQPSASSASTAAVAANTSSANTYPAGQCTGWGKSLAPWVNTWNGGQWAA 360  
DB 274 EPAALTA----PAAPSAGP---QTSYDSSNTYPVQCTGWGKSLAPWAGNNGGQWAY 326  
  
QY 361 SAAAGYRVGSGTSPAGAVAVWNDGGYGHVAVVTGVQ--GGQIQVCEANYAGNQSIGNYRGW 419  
DB 327 SAQAGYRTGSGTSPWGAIAVWNDGGYGHVAVVEVQSASSIRVWESNYGRQYIADHRGW 386  
  
QY 420 FPGSGVSYYIPN 431  
DB 387 FNPTGVTFIYPH 398

Search completed: February 10, 2006, 22:31:18  
Job time : 297.878 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:05:33 ; Search time 15.9234 Seconds  
(without alignments)  
607.053 Million cell updates/sec

Title: US-10-797-821-23

Perfect score: 113

Sequence: 1 TGAQTIKGKLYPKANGQQVKG 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_21.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113	100.0	22	2 AAR43696	Aar43696 GT subseq
2	113	100.0	22	2 AAW34159	Aaw34159 GTF antig
3	113	100.0	22	2 AAY43338	Aay43338 GTF antig
4	113	100.0	22	7 ADD93643	Add93643 Streptoco
5	113	100.0	22	9 ADX37266	Adx37266 Streptoco
6	113	100.0	24	2 AAW34164	Aaw34164 GTF antig
7	110	97.3	236	6 ABP56632	Abp56632 Cariogene
8	110	97.3	498	9 ADW43436	Adw43436 S. sobrin
9	110	97.3	1590	7 ADD93657	Add93657 Streptoco
10	110	97.3	1590	9 ADX37280	Adx37280 Streptoco
11	110	97.3	1592	2 AAR32925	Aar32925 Glucosylt
12	92	81.4	1499	7 ADC54806	Adc54806 Protein S
13	92	81.4	2835	5 ABB98574	Abb98574 Dextrans
14	92	81.4	2835	6 ABR55594	Abr55594 Amino aci
15	89	78.8	380	5 AAU79286	Aau79286 Streptoco
16	89	78.8	1475	5 AAU98031	Aau98031 S. mutans
17	89	78.8	1475	5 AAU98036	Aau98036 S. mutans
18	89	78.8	1475	5 AAU98037	Aau98037 S. mutans
19	89	78.8	1475	5 AAU98040	Aau98040 S. mutans
20	89	78.8	1475	5 AAU98035	Aau98035 S. mutans
21	89	78.8	1475	5 AAU98033	Aau98033 S. mutans
22	89	78.8	1475	5 AAU98034	Aau98034 S. mutans
23	89	78.8	1475	5 AAU98030	Aau98030 S. mutans
24	89	78.8	1475	5 AAU98039	Aau98039 S. mutans

25	89	78.8	1475	5 AAU98027	Aau98027 S. mutans
26	89	78.8	1475	5 AAU98032	Aau98032 S. mutans
27	89	78.8	1475	5 AAU98038	Aau98038 S. mutans
28	89	78.8	1475	7 ADD93654	Add93654 Streptoco
29	89	78.8	1475	9 ADX37277	Adx37277 Streptoco
30	89	78.8	1476	5 AAU79284	Aau79284 Streptoco
31	87	77.0	22	9 ADV68553	Adv68553 S. mutans
32	83	73.5	1518	7 ADD93660	Add93660 Streptoco
33	83	73.5	1518	9 ADX37283	Adx37283 Streptoco
34	82	72.6	1430	5 AAU98043	Aau98043 S. mutans
35	82	72.6	1430	5 AAU98044	Aau98044 S. mutans
36	82	72.6	1430	5 AAU98045	Aau98045 S. mutans
37	82	72.6	1430	5 AAU98042	Aau98042 S. mutans
38	82	72.6	1430	5 AAU98029	Aau98029 S. mutans
39	82	72.6	1430	5 AAU98041	Aau98041 S. mutans
40	82	72.6	1430	7 ADD93656	Add93656 Streptoco
41	82	72.6	1430	9 ADX37279	Adx37279 Streptoco
42	78	69.0	1365	7 ADD93659	Add93659 Streptoco
43	78	69.0	1365	9 ADX37282	Adx37282 Streptoco
44	74	65.5	1375	5 AAU98028	Aau98028 S. mutans
45	74	65.5	1375	5 AAU79288	Aau79288 Streptoco

ALIGNMENTS

RESULT 1  
AAR43696  
ID AAR43696 standard; peptide; 22 AA.  
XX AAR43696;  
AC AAR43696;  
XX  
DT 25-MAR-2003 (revised)  
DT 20-MAY-1994 (first entry)  
XX  
XX  
DE GT subsequence.  
XX  
KW GT; glucosyltransferase; vaccine; T-cell; B-cell; reaction;  
KW immunoresponse; peptidyl core matrix; dental caries; diptheria; tetanus;  
KW measles; polio.  
XX  
OS Synthetic.  
XX  
PN WO9322341-A1.  
XX  
XX  
PD 11-NOV-1993.  
XX  
PF 30-APR-1993; 93WO-US004094.  
XX  
PR 01-MAY-1992; 92US-00877295.  
XX  
PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.  
XX  
PI Taubman MA, Smith DJ;  
XX  
XX WPI; 1993-368721/46.  
PT Synthetic Glucosyl-transferase peptide(s) - provoke T- and B-cell  
PT reactions in mammals, and are effective against dental caries.  
XX  
PS Claim 3; Page 23; 38pp; English.  
XX  
CC The sequences (AAR43694-98) are subsequences from GT, they provoke T- and  
CC B-cell responses in mammals. The can be used to produce a vaccine  
CC comprising 2 of these sequences attached to a peptidyl core matrix. They  
CC are useful in producing T-cell responses to the GT enzyme causing a  
CC reduction of colonisation or accumulation of mutans streptococcal strains  
CC in mammals. The vaccines can be used in preventing dental caries.  
CC (Updated on 25-MAR-2003 to correct PN field.)  
XX  
SQ Sequence 22 AA;

Query Match 100.0%; Score 113; DB 2; Length 22;

Best Local Similarity 100.0%; Pred. No. 1.2e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIKGKLYFKANGQOVKG 22  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 TGAQTIKGKLYFKANGQOVKG 22

#### RESULT 2

ID AAW34159 standard; peptide; 22 AA.  
XX AAW34159;  
XX 25-MAR-2003 (revised)  
DT 18-FEB-1998 (first entry)  
XX  
XX GTF antigenic peptide #4.  
XX  
XX Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;  
KW surface domain; glucan-binding domain; mutans streptococcal strain;  
KW immune response; glucan-binding activity; dental caries prevention.  
XX

OS Streptococcus mutans.

XX US5686075-A.

PN 11-NOV-1997.

XX 30-APR-1993; 93US-00057162.

XX 01-MAY-1992; 92US-00877295.

XX (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.

XX Smith DJ, Taubman MA;

PI WPI; 1997-558089/51.

XX Immunogenic compositions containing streptococcal glucosyl:transferase  
PT peptide(s) - used for provoking immune response to streptococcal  
PT glucosyl:transferase for preventing dental caries.

XX Claim 2; Col 13; 11pp; English.

XX AAW34156-W34160 represent immunogenic fragments of the Streptomyces  
CC mutans glucosyltransferase (GTF) enzyme. AAW34157 and AAW34158 are from  
CC the catalytic domain of GTF, while AAW34160 is from the GTF surface  
CC domain. AAW34156 and AAW34159 are from the glucan-binding domain of GTF.  
CC These sequences can all be used in the immunogenic composition of the  
CC invention. The composition of the invention can alternatively comprise  
CC one or more of these sequences linked to a lysine core matrix (see  
CC AAW34161-W34165). A composition comprising one of these sequences can be  
CC administered to a mammal to raise an immune response, in a method for  
CC interfering with the enzymatic activity of streptococcal  
CC glucosyltransferase in a mammal. The immune response results in reduction  
CC of the colonisation or accumulation of mutans streptococcal strains in  
CC the mammal. Compositions containing AAW34156 specifically interfere with  
CC the glucan-binding activity of the streptococcal glucosyltransferase. The  
CC peptides can also be used in vaccines for preventing dental caries in  
CC mammals. (Updated on 25-MAR-2003 to correct PF field.)

XX Sequence 22 AA;

Query Match 100.0%; Score 113; DB 2; Length 22;

Best Local Similarity 100.0%; Pred. No. 1.2e-10;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIKGKLYFKANGQOVKG 22

Db 1 TGAQTIKGKLYFKANGQOVKG 22

#### RESULT 3

AAV43338  
ID AAY43338 standard; peptide; 22 AA.

XX AAY43338;

DT 26-JAN-2000 (first entry)

XX GTF antigenic epitope.

XX GTF; antigenic epitope; glucosyltransferase; immunogenic epitope; teeth;  
KW immunogenic composition; streptococcal glucan; antibody response;  
KW cariogenic bacteria; mutans streptococci; colonisation; caries;  
KW dental caries; immunisation; therapy.

OS Synthetic.

XX WO9952548-A2.

PN 21-OCT-1999.

XX 09-APR-1999; 99WO-US007828.

XX 10-APR-1998; 98US-0081315P.

XX (LEES/) LEES A.

PA (TAUB/) TAUBMAN M A.

PA (SMIT/) SMITH D J.

XX Lees A, Taubman MA, Smith DJ;

PI WPI; 1999-620289/53.

XX Immunogenic compositions for control of dental caries, based on

PT Streptococcus mutans components, particularly for vaccination of infants.

XX Claim 11; Page 44; 54pp; English.

XX This sequence represents a glucosyltransferase (GTF) immunogenic epitope,  
CC and can be used in the immunogenic composition of the invention. The  
CC immunogenic composition (A) comprises at least one streptococcal glucan  
CC (SG) and at least one component (I) covalently coupled, (in)directly, to  
CC SG. (A) elicits antibodies (Ab) to both SG and (I). (A), and related  
CC compositions, are used to induce an antibody response to cariogenic  
CC bacteria, collectively called "mutans streptococci" (i.e. any of  
CC Streptococcus mutans, S. rattus, S. cricetus or S. sobrinus), especially  
CC to prevent colonisation, to reduce numbers of bacteria or to reduce  
CC incidence of caries, in humans or other animals. The composition can  
CC therefore be used for the treatment or prevention of caries (particularly  
CC dental caries). Abs raised against (A) can be used: (a) to detect SG, and  
CC (b) for passive immunisation, e.g. applied directly to the surface of  
CC teeth. SG is a poorly immunogenic type 2 T-cell independent antigen but  
CC when coupled to (I) it produces a conjugate which induces significant  
CC immunity to both components, resulting in immunological memory and long-  
CC lasting antibody production against SG, in adults and children

XX Sequence 22 AA;

Query Match 100.0%; Score 113; DB 2; Length 22;

Best Local Similarity 100.0%; Pred. No. 1.2e-10;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIKGKLYFKANGQOVKG 22

Db 1 TGAQTIKGKLYFKANGQOVKG 22

#### RESULT 4

ADD93643

ID ADD93643 standard; peptide; 22 AA.

XX ADD93643;

AC ADD93643;

XX

```

DT 29-JAN-2004 (first entry)
XX Streptococcus sobrinus glucosyltransferase-B peptide fragment.
DE Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
KW Streptococcus sobrinus.
XX
OS Streptococcus sobrinus.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
XX 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
XX
PA (FORSYTH INST.
XX
PI Smith DJ, Taubman MA;
XX
XX WPI; 2003-845091/78.
XX
XX Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
XX Claim 19; Page 17; 49pp; English.
XX
CC The present sequence is that of a glucan-binding peptide comprising amino
CC acid residues 1303-1324 of Streptococcus sobrinus glucosyltransferase.
CC The peptide can be used in immunogenic compositions and subunit vaccines
CC for dental caries. These compositions comprise a major histocompatibility
CC complex (MHC) class II protein-binding peptide from Streptococcus mutans
CC glucan binding protein-B (GbpB) covalently linked with a peptide subunit
CC (such as the present peptide) of a streptococcal glucosyltransferase. The
CC compositions are used in a claimed method of eliciting production of an
CC antibody in a mammal. Diepitopic or multi-epitopic polypeptides can be
CC prepared synthetically or by recombinant DNA technology. Antibodies
CC raised against MHC class II binding fragments of GbpB can be used in
CC passive immunisation.
XX
XX Sequence 22 AA;
XX
Query Match 100.0%; Score 113; DB 7; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIGQKLYFKANGQQVKG 22
Db |||||
1 TGAQTIGQKLYFKANGQQVKG 22

RESULT 5
ADX37266
ID ADX37266 standard; peptide; 22 AA.
XX
AC ADX37266;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B peptide #23.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
XX 09-MAR-2004; 2004US-00797821.
XX

```

---

```

XX 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
XX (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
XX Smith DJ, Taubman MA;
PI WPI; 2005-151644/16.
XX
XX New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
XX Claim 10; SEQ ID NO 23; 73pp; English.
XX
XX The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB-derived peptide of the invention.
XX
XX Sequence 22 AA;
XX
Query Match 100.0%; Score 113; DB 9; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.2e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIGQKLYFKANGQQVKG 22
Db |||||
1 TGAQTIGQKLYFKANGQQVKG 22

RESULT 6
AAW34164
ID AAW34164 standard; peptide; 24 AA.
XX
AC AAW34164;
XX
DT 25-MAR-2003 (revised)
DT 18-FEB-1998 (first entry)
XX
DE GTF antigenic peptide #4 linked to polylysine core.
XX
KW Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;
KW surface domain; glucan-binding domain; mutans streptococcal strain;
KW immune response; glucan-binding activity; dental caries prevention.
XX
OS Synthetic.
OS Streptococcus mutans.
XX
XX Key Location/Qualifiers
FT Peptide 1..22
FT /label= GTF antigenic peptide #4 (see AAW34159)
FT /note= "attached to the dendritic polylysine core via the
FT alpha-amino group of Lys(23); a second copy of the
FT antigenic 22-mer is linked to Lys(23) via the omega amino
FT group"
FT Modified-site 23
FT /note= "Lys(23) is linked to one copy of the antigenic
FT peptide through the alpha-amino group, and to a second
FT copy of the peptide (not shown) via the omega amino
FT group"
FT Modified-site 24
FT /note= "the alpha amino acid group of Lys(24) forms a
FT peptide linkage with the carboxyl amino group of Lys(23);
FT

```

FT the omega amino group of Lys(24) forms a peptide bond  
 FT with a second Lys residue analogous to Lys(23)"  
 XX  
 PN US5686075-A.  
 XX  
 PD 11-NOV-1997.  
 XX  
 PF 30-APR-1993; 93US-00057162.  
 XX  
 PR 01-MAY-1992; 92US-00877295.  
 XX  
 PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX  
 DR WPI; 1997-558089/51.  
 XX  
 PT Immunogenic compositions containing streptococcal glucosyl:transferase  
 PT peptide(s) - used for provoking immune response to streptococcal  
 PT glucosyl:transferase for preventing dental caries.  
 XX  
 PS Claim 12; Col; 11pp; English.  
 XX  
 CC AAW34161-W34165 represent the Streptomyces mutans glycosyltransferase  
 CC (GTF) enzyme immunogenic fragments shown in AAW34156-W34160 linked to a  
 CC polylysine core. AAW34157 and AAW34158 are from the catalytic domain of  
 CC GTF, while AAW34160 is from the GTF surface domain. AAW34156 and AAW34159  
 CC are from the glucan-binding domain of GTF. These sequences, and the  
 CC immunogenic fragments shown in AAW34156-W34160 can all be used in the  
 CC immunogenic composition of the invention. A composition comprising one of  
 CC these sequences can be administered to a mammal to raise an immune  
 CC response, in a method for interfering with the enzymatic activity of  
 CC streptococcal glucosyltransferase in a mammal. The immune response  
 CC results in reduction of the colonisation or accumulation of mutants  
 CC streptococcal strains in the mammal. Compositions containing AAW34156  
 CC specifically interfere with the glucan-binding activity of the  
 CC streptococcal glucosyltransferase. The peptides can also be used in  
 CC vaccines for preventing dental caries in mammals. (Updated on 25-MAR-2003  
 CC to correct PF field.)  
 XX  
 SQ Sequence 24 AA;

Query Match 100.0%; Score 113; DB 2; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-10;  
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYPKANGQQVKG 22  
 |||||  
 DB 1 TGAQTIKGQKLYPKANGQQVKG 22

RESULT 7  
 ABP56632  
 ID ABP56632 standard; protein; 236 AA.

AC ABP56632;  
 XX  
 DT 24-MAR-2003 (first entry)  
 XX

DE Cariogenetic streptococci dextran-combining domain SEQ ID NO:1.  
 XX  
 XX Cariogenetic streptococci; dextran-combining domain; purification.

OS Streptococcus sp.  
 OS Synthetic.  
 PN JP2002233370-A.  
 XX

PD 20-AUG-2002.  
 XX

PF 05-FEB-2001; 2001JP-00028637.  
 XX  
 PR 05-FEB-2001; 2001JP-00028637.

XX  
 PA (DOKU-) DOKURITSU GYOSEI HOJIN SANGYO GIJUTSU SO.  
 XX  
 DR WPI; 2003-096386/09.  
 DR N-PSDB; ABZ22469.  
 XX  
 PT Purification of a protein useful in pharmacology and medical science,  
 PT comprises eluting the protein in a solution containing dextran.  
 XX  
 PS Claim 1; Page 7-8; 10pp; Japanese.  
 XX  
 CC The present invention describes the purification of a protein comprising  
 CC mixing an adsorbent containing a crosslinking dextran with a solution  
 CC containing a fused protein. The protein molecule to be purified is fused  
 CC with a dextran-combining domain of a sequence of 236 amino acids (see  
 CC ABP56632, SI), or a dextran-combining domain having an amino acid  
 CC sequence wherein at least one amino acid is deleted, replaced or added in  
 CC (SI), and has dextran-combining activity. The adsorbent taken out of  
 CC solution is then mixed with a solution containing dextran and the fused  
 CC protein is eluted in the solution containing dextran. The method is used  
 CC for the purification of a protein  
 XX  
 SQ Sequence 236 AA;

Query Match 97.3%; Score 110; DB 6; Length 236;  
 Best Local Similarity 95.5%; Pred. No. 5.6e-09;  
 Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYPKANGQQVKG 22  
 |||||  
 DB 205 TGAQTIRGQKLYPKANGQQVKG 226

RESULT 8  
 ADW43436  
 ID ADW43436 standard; protein; 498 AA.

XX ADW43436;  
 XX  
 DT 24-MAR-2005 (first entry)  
 XX

DE S. sobrinus fusion protein associated protein, SEQ ID NO:2.

KW Fusion protein; microbial cell-surface layer protein; glucan synthetase;  
 KW recombinant DNA; antibody production; dental caries; tooth; phagocytosis;  
 KW antibacterial.

OS Streptococcus sobrinus.  
 XX  
 PN WO2004099418-A1.  
 XX

PD 18-NOV-2004.

PF 12-MAY-2004; 2004WO-JP006393.

PR 12-MAY-2003; 2003JP-00132892.

XX (UYNI-) UNIV NIPPON.

XX Yamashita Y, Kawato T;

DR WPI; 2005-131991/14.

DR N-PSDB; ADW43438.

XX Fusion protein useful for preventing dental caries, comprises alanine-  
 PT repeat region of high molecular weight microbial cell-surface layer  
 PT protein of Streptococcus sobrinus, and glucan-binding region of glucan  
 PT synthetase.

XX Claim 3; SEQ ID NO 2; 24pp; Japanese.

XX The invention relates to a fusion protein comprising an alanine-repeat  
 CC region of a high molecular weight microbial cell-surface layer protein of

CC Streptococcus sobrinus, and a glucan-binding region of glucan synthetase.  
 CC Also described is a recombinant DNA encoding the fusion protein of the  
 CC invention, an expression vector containing the recombinant DNA, a  
 CC transformed host containing the expression vector, and an antibody  
 CC formulation containing the antibody with respect to the fusion protein.  
 CC The fusion protein and the antibody formulation of the invention are  
 CC useful for preventing dental caries caused by *S. sobrinus*. The fusion  
 CC protein allows the inhibition of adhesion of microorganisms such as *S.*  
 CC *sobrinus* on the surface of teeth, and induces phagocytosis of microbial  
 CC cells, and thus prevents dental caries in an individual. This sequence  
 CC represents a protein relating to the present invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at fcp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 498 AA;

Query Match 97.3%; Score 110; DB 9; Length 498;  
 Best Local Similarity 95.5%; Pred. No. 1.3e-08;  
 Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIRGQKLYFKANGQQVKG 22  
 |||||:|||||  
 Db 204 TGAQTIRGQKLYFKANGQQVKG 225

RESULT 9  
 ADD93657  
 ID ADD93657 standard; protein; 1590 AA.

XX AC ADD93657;

XX DT 29-JAN-2004 (first entry)

XX DE Streptococcus sobrinus glucosyltransferase-I.

XX KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.

XX OS Streptococcus sobrinus.

XX PN W02003075845-A2.

XX PD 18-SEP-2003.

XX PF 07-MAR-2003; 2003WO-US006962.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX PA (FORS-) FORSYTH INST.

XX PI Smith DJ, Taubman MA;

XX PS WPI; 2003-845091/78.

XX CC Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.

XX PS Claim 16; Page 14; 49pp; English.

XX CC The present sequence is the protein sequence of Streptococcus sobrinus  
 CC glucosyltransferase-I (GTF-I). Peptide fragments of GTF-I, especially  
 CC from the catalytic domain of the polypeptide, can be used in immunogenic  
 CC compositions and subunit vaccines for dental caries. These compositions  
 CC comprise a major histocompatibility complex (MHC) class II protein-  
 CC binding peptide from *S. mutants* glucan binding protein-B (GbpB)  
 CC covalently linked with a peptide fragment of a streptococcal  
 CC glucosyltransferase. The compositions are used in a claimed method of  
 CC eliciting production of an antibody in a mammal. Diepitopic or  
 CC multiepitopic polypeptides can be prepared synthetically or by  
 CC recombinant DNA technology. Antibodies raised against MHC class II  
 CC binding fragments of GbpB can be used in passive immunisation.

XX SQ Sequence 1590 AA;

Query Match 97.3%; Score 110; DB 7; Length 1590;  
 Best Local Similarity 95.5%; Pred. No. 5.1e-08;  
 Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIRGQKLYFKANGQQVKG 22  
 |||||:|||||  
 Db 1296 TGAQTIRGQKLYFKANGQQVKG 1317

RESULT 10

ADK37280  
 ID ADK37280 standard; protein; 1590 AA.

XX AC ADK37280;

XX DT 21-APR-2005 (first entry)

XX DE Streptococcus sobrinus glucan binding protein B.

XX KW immunogenicity; immune stimulation; glucan binding protein-B;  
 KM microparticle; major histocompatibility complex; tooth disease.

XX OS Streptococcus sobrinus.

XX PN US2005031633-A1.

XX PD 10-FEB-2005.

XX PF 09-MAR-2004; 2004US-00797821.

XX PR 13-APR-1998; 98US-0081550P.

XX PR 08-JAN-1999; 99US-0115142P.

XX PR 12-APR-1999; 99US-00290049.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX PR 07-MAR-2003; 2003US-00383930.

XX PA (SMIT/) SMITH D J.

XX PA (TAUB/) TAUBMAN M A.

XX PI Smith DJ, Taubman MA;

XX DR WPI; 2005-151644/16.

XX CC New composition comprising a fragment of a glucan binding protein-B  
 PT (GbpB) that binds to MHC class II protein, and a biocompatible  
 PT microparticle, useful for producing an antibody (claimed) for immunizing  
 PT mammals against dental caries.

XX PS Claim 7; SEQ ID NO 37; 73pp; English.

XX CC The invention relates to a composition comprising a fragment of a glucan  
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II  
 CC protein. The composition is useful for producing an antibody for  
 CC immunizing mammals against dental caries. This sequence corresponds to a  
 CC Streptococcus sobrinus GbpB protein of the invention.

XX SQ Sequence 1590 AA;

Query Match 97.3%; Score 110; DB 9; Length 1590;  
 Best Local Similarity 95.5%; Pred. No. 5.1e-08;  
 Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIRGQKLYFKANGQQVKG 22  
 |||||:|||||  
 Db 1296 TGAQTIRGQKLYFKANGQQVKG 1317

RESULT 11

```

AAR32925
ID AAR32925 standard; protein; 1592 AA.
XX
AC AAR32925;
XX
XX 28-JUN-1993 (first entry)
DE
DE Glucosyltransferase I.
XX
XX GT-1; Streptococcus; dental; caries.
XX
XX Streptococcus sobrinus.
OS
XX
XX JP05023188-A.
PN
XX
XX 02-FEB-1993.
PD
XX
XX 25-JUL-1991; 91JP-00186592.
PF
XX
XX 25-JUL-1991; 91JP-00186592.
PR
XX
XX (KATO/) KATO K.
PA
XX (FUKU/) FUKUI I.
PA
XX
XX WPI; 1993-079449/10.
DR
XX
XX N-PSDB; AAQ37760.
DR
XX
XX DNA sequence Glucosyltransferase-I - comprises Streptococcus sobrinus
PT DNA sequence with at least one nucleotide added or deleted.
PT
XX
XX Claim 13; Page 15; 29pp; Japanese.
PS
XX
XX The DNA sequence from Streptococcus sobrinus strain 6715 encodes
CC Glucosyltransferase-I (and mutants). The DNA was obcd. by treating S.
CC sobrinus 6715 with mutanolysin, extracting the chromosomal DNA, partially
CC digesting with Sau3AI and fractionating on agarose gel. The 3-5 kbp
CC fragment was ligated into pUC18 and E. coli JM109 transformed with it. A
CC GT-1 expressing clone was isolated and sequenced. The clone may be used
CC in the development of a drug for dental caries
XX
XX
XX Sequence 1592 AA;
SQ
Query Match 97.3%; Score 110; DB 2; Length 1592;
Best Local Similarity 95.5%; Pred. No. 5.1e-08;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22
Db 1298 TGAQTIKGQKLYFKANGQQVKG 1319

RESULT 12
ADC54806
ID ADC54806 standard; protein; 1499 AA.
XX
XX ADC54806;
AC
XX
XX 18-DEC-2003 (first entry)
DT
XX
XX Protein Seq ID11 related to L mesenteroides dextran sucrose protein.
DE
XX
XX Dextran sucrose; active centre zone; glucan; polysaccharide; dextran;
KW D-glucose; starch; cellulose; glucan manufacture; transduction;
KW enzyme-reaction product.
XX
XX Unidentified.
OS
XX
XX JP2003111590-A.
PN
XX
XX 15-APR-2003.
XX
XX 03-OCT-2001; 2001JP-00307067.
PF
XX
XX
XX

```

---

```

PR 03-OCT-2001; 2001JP-00307067.
XX
XX (DOKU-) DOKURITSU GYOSEI HOJIN SHOKUHIN SOGO KEN.
PA
XX
XX WPI; 2003-735670/70.
DR
XX P-PSDB; ADC54814.
DR
XX
XX Novel modified dextran sucrose which exchanges one site of active center
PT zone of dextran sucrose for active center zone of different types of
PT dextran sucrose, useful for manufacturing glucan.
PT
XX
XX Example 1; SEQ ID NO 11; 28pp; Japanese.
PS
XX
XX This invention relates to a modified dextran sucrose (DS) exchanging one
XX site of the active centre zone of a dextran sucrose for the active centre
CC zone of different types of dextran sucrose. Glucan is a polysaccharide
CC (for example dextran) which uses D-glucoses, such as a starch and a
CC cellulose, as a structural unit. The modified enzyme of the invention is
CC useful in the manufacture of glucan. The selection of the active centre
CC area of DS which carries out transduction, enables changes in structure
CC and character of an enzyme-reaction product and their application to
CC various uses. The present sequence is that of a protein which is related
CC to the Leuconostoc mesenteroides dextran sucrose protein and which was
CC used during the exemplification of the invention.
XX
XX Sequence 1499 AA;
SQ
Query Match 81.4%; Score 92; DB 7; Length 1499;
Best Local Similarity 77.3%; Pred. No. 3.4e-05;
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22
Db 1325 TGAQTIKGQKLYFKANGQQVKG 1346

RESULT 13
AB998574
ID AB998574 standard; protein; 2835 AA.
XX
XX AB998574;
AC
XX
XX 29-AUG-2003 (revised)
DT
XX 14-JAN-2003 (first entry)
DT
XX
XX Dextran saccharase, DSRE.
DE
XX
XX Dextran saccharase; enzyme; cytostatic; dermatological; antiseborrheic;
KW DSR-E; glycosyl transferase; dextran; prebiotic; pharmaceutical;
KW microflora regulation; intestinal transit; mineral assimilation;
KW colon cancer; acne; dandruff; body odour.
XX
XX Leuconostoc mesenteroides; NRRL B-1299.
OS
XX
XX Key Location/Qualifiers
FH Peptide 1..40
FT /label= Signal_peptide
FT Protein 41..2835
FT /label= Mature_protein
FT Peptide 423..439
FT /label= SEQ_ID_6
FT /note= "This sequence is specifically claimed in Claim 7"
FT Peptide 478..501
FT /label= SEQ_ID_7
FT /note= "This sequence is specifically claimed in Claim 7"
FT Peptide 519..539
FT /label= SEQ_ID_8
FT /note= "This sequence is specifically claimed in Claim 7"
FT Peptide 560..571
FT /label= SEQ_ID_9
FT /note= "This sequence is specifically claimed in Claim 7"
FT Peptide 631..645
FT /label= SEQ_ID_10
FT

```

FT Peptide /note= "This sequence is specifically claimed in Claim 7"  
 FT 1014. .1021  
 FT /label= SEQ\_ID\_11  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 1981. .1142  
 FT Domain /note= "Catalytic domain, SEQ ID 1. This sequence is  
 FT specifically claimed in Claim 4"  
 FT 2120. .2138  
 FT Peptide /label= SEQ\_ID\_12  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 2161. .2184  
 FT Peptide /label= SEQ\_ID\_13  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 2202. .2214  
 FT Peptide /label= SEQ\_ID\_14  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 2243. .2250  
 FT Peptide /label= SEQ\_ID\_15  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 2315. .2322  
 FT Peptide /label= SEQ\_ID\_16  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT 2689. .2696  
 FT Peptide /label= SEQ\_ID\_17  
 FT /note= "This sequence is specifically claimed in Claim 7"  
 FT  
 XX FR2822163-A1.  
 PN  
 XX  
 XX 20-SEP-2002.  
 PD  
 XX  
 XX 19-DEC-2001; 2001FR-00016495.  
 PF  
 XX  
 XX 16-MAR-2001; 2001FR-00003631.  
 PR  
 XX  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 PA  
 XX  
 XX Bozonnet SAM, Renaud SMMC, Willemot RML, Monsan PEF;  
 PI  
 XX  
 XX WPI; 2002-715213/78.  
 DR  
 XX N-PSDB; ABQ80961, ABQ80962.  
 DR  
 XX  
 XX New glycosyl transferase enzymes, containing glucan bonding and catalytic  
 PT domains and producing alpha-(1-2) branched dextrans, useful in probiotic,  
 PT pharmaceutical or cosmetic compositions.  
 XX  
 XX Claim 6; Page 65-74; 82pp; French.  
 PS  
 XX  
 XX The present sequence is a novel dextran saccharase, DSR-E, from  
 CC Leuconostoc mesenteroides NRRL B-1299. The saccharase has glycosyl  
 CC transferase activity suitable for producing dextrans having alpha(1-2)  
 CC branches from sucrose, p-nitrophenyl-alpha-D-glucopyranoside, alpha-  
 CC fluoroglucose, alpha-D-glucopyranoside-alpha-D-sorbofuranoside or alpha-D-  
 CC galactopyranosyl-sucrose. The dextran saccharase is useful in probiotic,  
 CC pharmaceutical or cosmetic compositions. The dextrans and related  
 CC compounds having alpha(1-2) bonds, produced using DSR-E, may be involved  
 CC in signalling/cellular recognition processes in vivo (specifically in  
 CC regulation of microflora in the intestines or on the skin); and are  
 CC potentially useful for improving intestinal transit, increasing  
 CC assimilation of minerals (e.g. calcium and/or magnesium), preventing  
 CC cancer of the colon and combating skin problems such as acne, dandruff  
 CC and body odour. (Updated on 29-AUG-2003 to standardise OS field)  
 XX  
 XX Sequence 2835 AA;  
 SQ  
 Query Match 81.4%; Score 92; DB 5; Length 2835;  
 Best Local Similarity 81.8%; Pred. No. 7.1e-05;  
 Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 TGAQTIKGQKLYFKANGQQVKG 22  
 DB 1644 TGAQIIINGQNLVFKADGQQVKG 1665

RESULT 14  
 ABR55594  
 ID ABR55594 standard; protein; 2835 AA.  
 XX  
 AC ABR55594;  
 XX  
 DT 11-AUG-2003 (first entry)  
 XX  
 DE Amino acid sequence of a glycosyltransferase designated DSR-D.  
 XX  
 KW Glycosyltransferase; DSR-D; enzyme; dextran; sucrose;  
 KW alpha-D-fluoroglucose; p-nitrophenol alpha-D-glucopyranoside;  
 KW alpha-D-glucopyranoside-alpha-D-sorbofuranoside; oligosaccharide;  
 KW oligoside; 4-O-alpha-D-galactopyranosylsucrose; weight gain;  
 KW colon cancer; skin disorder; acne; dandruff; body odour;  
 KW intestinal transit; calcium; magnesium.  
 XX  
 OS Leuconostoc mesenteroides.  
 XX  
 XX FR2822162-A1.  
 PN  
 XX  
 XX 20-SEP-2002.  
 PD  
 XX  
 XX 16-MAR-2001; 2001FR-00003631.  
 PF  
 XX  
 XX 16-MAR-2001; 2001FR-00003631.  
 PR  
 XX  
 XX (NASC-) INST NAT SCI APPLIQUEES TOULOUSE.  
 PA  
 XX  
 XX (CNRS ) CNRS CENT NAT RECH SCI.  
 XX  
 PI Monsan P;  
 XX  
 XX WPI; 2003-432740/41.  
 DR  
 XX N-PSDB; ACC70331.  
 DR  
 XX  
 XX New polypeptide with glycosyltransferase activity, useful for producing  
 PT dextrans with alpha 1-2 branches, e.g. used as prebiotics.  
 PT  
 XX  
 XX Claim 6; Fig 8; 103pp; French.  
 PS  
 XX  
 XX The present sequence represents a glycosyltransferase, designated DSR-D.  
 CC This enzyme produces dextrans having alpha(1-2) branches from sucrose,  
 CC alpha-D-fluoroglucose, p-nitrophenol alpha-D-glucopyranoside, alpha-D-  
 CC glucopyranoside-alpha-D-sorbofuranoside or 4-O-alpha-D-  
 CC galactopyranosylsucrose. The polypeptide is used to prepare  
 CC oligosaccharides and oligosides for prebiotic, pharmaceutical, diagnostic  
 CC and cosmetic applications, e.g. to improve health and weight gain in  
 CC animals; prevent cancer of the colon; and treat skin disorders such as  
 CC acne, dandruff and body odour. The oligosaccharides and oligosides  
 CC produced by DSR-D improve intestinal transit and assimilation of calcium,  
 CC magnesium and other minerals  
 XX  
 XX Sequence 2835 AA;  
 SQ  
 Query Match 81.4%; Score 92; DB 6; Length 2835;  
 Best Local Similarity 81.8%; Pred. No. 7.1e-05;  
 Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 TGAQTIKGQKLYFKANGQQVKG 22  
 DB 1644 TGAQIIINGQNLVFKADGQQVKG 1665

RESULT 15  
 AAU79286  
 ID AAU79286 standard; protein; 380 AA.  
 XX  
 AC AAU79286;  
 XX  
 DT 13-AUG-2002 (first entry)  
 XX  
 DE Streptococcus mutans monoclonal antibody-related protein #3.  
 XX

KW Antibody; dental caries; water insoluble glucan synthetase; anti-caries;  
KW glucosyl transferase-B; immunotherapy.  
XX  
OS Streptococcus mutans.  
XX  
PN JP2002114709-A.  
XX  
XX 16-APR-2002.  
XX  
PD  
XX  
PF 04-OCT-2000; 2000JP-00304889.  
XX  
PR 04-OCT-2000; 2000JP-00304889.  
XX  
XX (UYNI-) UNIV NIPPON.  
PA  
XX WPI; 2002-448885/48.  
DR  
XX  
XX Anti-caries agent composed of a monoclonal antibody against an inhibitory  
PT enzyme against water insoluble glucan synthetase of glucosyl transferase-  
PT B (GTF-B) of Streptococcus mutans.  
XX  
XX Claim 5; Page 19-20; 28pp; Japanese.  
XX  
CC The invention relates to a monoclonal antibody against dental caries and  
CC an anti-caries agent composed of a monoclonal antibody produced by  
CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)  
CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having  
CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
CC transferase-B. The monoclonal antibody specifically inhibits water  
CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl  
CC transferase-B and is used in the immunotherapy of dental caries. This  
CC sequence represents a Streptococcus mutans monoclonal antibody-related  
CC protein  
XX  
SQ Sequence 380 AA;

Query Match 78.8%; Score 89; DB 5; Length 380;  
Best Local Similarity 77.3%; Pred. No. 2.1e-05;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQVKG 22  
||:|||||:|||||  
Db 140 TGSQTINGQHLVFRANGVQVKG 161

Search completed: February 10, 2006, 22:19:41  
Job time : 16.9234 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:20:08 ; Search time 2.59476 Seconds  
(without alignments)  
815.787 Million cell updates/sec

Title: US-10-797-821-23  
Perfect score: 113  
Sequence: 1 TGAQTIGQKLYFKANGQQVKG 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	97.3	1592	2 A38175	glucosyltransferas
2	89	78.8	1475	2 B33135	gtfB protein precu
3	87	77.0	1599	2 S22737	glucosyltransferas
4	83	73.5	1518	2 A44811	glucosyltransferas
5	82	72.6	1431	2 A45866	dextranucrase (EC
6	78	69.0	1290	2 JC5473	dextranucrase (EC
7	78	69.0	1365	2 A41483	glucosyltransferas
8	74	65.5	1375	2 JT0345	dextranucrase (EC
9	74	65.5	1449	2 T30857	glucosyltransferas
10	74	65.5	1449	2 T30552	glucosyltransferas
11	71	62.8	1577	2 T30858	glucosyltransferas
12	67.5	59.7	329	2 A55221	dextranase inhibit
13	66	58.4	563	2 A37184	glucan-binding pro
14	65	57.5	1508	2 T31098	probable dextranu
15	58	51.3	831	2 T00323	chitinase (EC 3.2
16	58	51.3	2817	2 B97033	uncharacterized pr
17	52	46.0	299	2 T15928	hypothetical prote
18	52	46.0	2178	2 S55805	alpha-toxin - Clos
19	52	46.0	2710	2 A37052	toxin A - Clostrid
20	51	45.1	648	2 S10869	enterotoxin A - Cl
21	51	45.1	696	2 A29635	transcription fact
22	51	45.1	788	2 J50747	regulatory protein
23	47	41.6	146	2 H81851	conserved hypochet
24	47	41.6	194	2 G64019	hypothetical prote
25	45	39.8	480	2 G71664	UDP-n-acetylmutamo
26	45	39.8	407	2 T39658	probable mitochond
27	45	39.8	466	2 G84476	probable retroelem
28	45	39.8	589	2 F97735	hypothetical prote
29	44	38.9	305	2 B75474	probable chloride

30	44	38.9	592	2 A81888	probable surface f
31	44	38.9	908	2 AE2254	hypothetical prote
32	44	38.9	1301	2 S18118	alpha-amylase - Al
33	44	38.9	1983	2 G86643	hypothetical prote
34	43.5	38.5	417	2 C71212	probable dihydroor
35	43.5	38.5	441	2 T01000	hypothetical prote
36	43	38.1	42	2 F97915	choline binding pr
37	43	38.1	128	1 CCRFCX	cytochrome c' - Rh
38	43	38.1	275	2 A82188	conserved hypochet
39	43	38.1	285	2 C95045	choline binding pr
40	43	38.1	340	2 G95043	choline binding pr
41	43	38.1	393	2 AB2664	conserved hypochet
42	43	38.1	401	2 A97446	hypothetical prote
43	43	38.1	500	2 S16788	hypothetical prote
44	43	38.1	591	2 C84220	propionyl-CoA carb
45	43	38.1	928	2 C81265	probable lipoprote

ALIGNMENTS

RESULT 1

A38175  
glucosyltransferase precursor - Streptococcus sobrinus  
C;Species: Streptococcus sobrinus  
C;Date: 28-Aug-1992 #sequence\_revision 28-Aug-1992 #text\_change 15-Mar-2004  
C;Accession: A38175  
R;Abn, H.; Mtsamura, T.; Kodama, T.; Ohta, H.; Fukui, K.; Kagawa, H.  
J. Bacteriol. 173, 989-996, 1991  
A;Title: Peptide sequences for sucrose splitting and glucan binding within Streptococcus  
A;Reference number: A38175; MUID:91123227; PMID:1704006  
A;Accession: A38175  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1592 <AB>  
A;Cross-references: UNIPARC:UPI0000128CB2; GB:D90213; NID:g217032; PIDN:BAA14241.1; PID:1093-1112/Domain: cpl repeat homology <CP1>  
F;1222-1241/Domain: cpl repeat homology <CP2>  
F;1287-1306/Domain: cpl repeat homology <CP3>  
F;1330-1351/Domain: cpl repeat homology <CP4>  
F;1352-1371/Domain: cpl repeat homology <CP5>  
F;1402-1420/Domain: cpl repeat homology <CP6>  
F;1465-1484/Domain: cpl repeat homology <CP7>  
F;1513-1532/Domain: cpl repeat homology <CP8>

Query Match 97.3%; Score 110; DB 2; Length 1592;  
Best Local Similarity 95.5%; Pred. No. 7.9e-09;  
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGAQTIGQKLYFKANGQQVKG 22  
|||:|||||  
Db 1298 TGAQTIRGQKLYFKANGQQVKG 1319

RESULT 2

B33135  
gtfB protein precursor - Streptococcus mutans  
C;Species: Streptococcus mutans  
C;Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 09-Jul-2004  
C;Accession: B33135; A33128  
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A;Reference number: A33135; MUID:87308013; PMID:3040685  
A;Accession: B33135  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1475 <SH1>  
A;Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:g153639; PIDN:1-1475/Domain: cpl repeat homology <CP1>  
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
submitted to the Protein Sequence Database, September 1990  
A;Reference number: A33128  
A;Accession: A33128







GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:07:53 ; Search time 15.1028 Seconds  
(without alignments)  
1027.730 Million cell updates/sec

Title: US-10-797-821-23  
Perfect score: 113  
Sequence: 1 TGAQTIGKQKLYFKANGQQVK 22

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	113	100.0	1597	1	GTF1_STRDO	P11001 streptococc
2	110	97.3	1590	2	Q59883_9STRE	Q59883 streptococc
3	110	97.3	1592	1	GTF2_STRDO	P27470 streptococc
4	98	86.7	1590	2	Q55263_9STRE	Q55263 streptococc
5	92	81.4	1454	2	Q69A94_LEUME	Q69A94 leuconostoc
6	92	81.4	2835	2	Q8G9Q2_LEUME	Q8G9Q2 leuconostoc
7	89	78.8	1476	1	GTFB_STRMU	P08987 streptococc
8	87	77.0	1599	2	Q00599_STRSL	Q00599 streptococc
9	83	73.5	1518	2	Q00600_STRSL	Q00600 streptococc
10	82	72.6	1462	1	GTFD_STRMU	P49331 streptococc
11	81	71.7	591	2	Q8VUH3_STRMU	Q8VUH3 streptococc
12	81	71.7	1338	2	Q9WJ14_9STRE	Q9WJ14 streptococc
13	79.5	70.4	1512	2	Q9WJ15_9STRE	Q9WJ15 streptococc
14	78	69.0	1290	2	Q48756_LEUME	Q48756 leuconostoc
15	78	69.0	1365	1	GTF5_STRDO	P29336 streptococc
16	77	68.1	1455	1	GTF6_STRMU	P13470 streptococc
17	74.5	65.9	1506	2	Q56CX8_9STRE	Q56CX8 streptococc
18	74	65.5	1449	2	Q68542_STRSL	Q68542 streptococc
19	74	65.5	1449	2	Q55264_STRSL	Q55264 streptococc
20	73	64.6	726	2	Q8DUW9_STRMU	Q8DUW9 streptococc
21	72	63.7	565	2	Q8DRV2_STRMU	Q8DRV2 streptococc
22	71	62.8	1577	2	Q55265_STRSL	Q55265 streptococc
23	67.5	59.7	330	2	Q55228_9STRE	Q55228 streptococc
24	67	59.3	1577	2	Q54178_STRGN	Q54178 streptococc
25	66	58.4	563	2	Q54447_STRMU	Q54447 streptococc
26	65	57.5	1330	2	Q84CN4_LEUME	Q84CN4 leuconostoc
27	65	57.5	1463	2	Q5SBM6_LACPE	Q5sbm6 lactobacill
28	65	57.5	1477	2	Q9L466_LEUME	Q9L466 leuconostoc
29	65	57.5	1508	2	Q9E2H5_LEUME	Q9e2h5 leuconostoc
30	65	57.5	1508	2	O52224_LEUME	O52224 leuconostoc
31	65	57.5	1561	2	Q5SBM8_9LACO	Q5sbm8 lactobacill

32	63	55.8	1522	2	Q6TXV4_LEUME	Q6txv4 leuconostoc
33	63	55.8	1527	2	Q8KRE1_LEUME	Q8krel leuconostoc
34	63	55.8	1527	2	Q9ZAR4_LEUME	Q9zar4 leuconostoc
35	62	54.9	1575	2	Q9LCH3_STROR	Q9lch3 streptococc
36	61	54.0	2057	2	Q9RE05_LEUME	Q9re05 leuconostoc
37	58	51.3	831	2	O50076_9CLOT	O50076 clostridium
38	58	51.3	1595	2	Q5SBM3_LACSK	Q5sbm3 lactobacill
39	58	51.3	2817	2	Q97K42_CLOAB	Q97k42 clostridium
40	53	46.9	1187	2	O589G5_CHICK	O589g5 gallus gall
41	52.5	46.5	1554	2	Q8KZL5_9STRE	Q8kz15 streptococc
42	52	46.0	299	1	YQO2_CAEEL	Q9529 caenorhabdi
43	52	46.0	929	2	O6XD70_CLODI	O6xd70 clostridium
44	52	46.0	975	2	Q798V2_CLONO	Q798v2 clostridium
45	52	46.0	2178	2	Q46149_CLONO	Q46149 clostridium

ALIGNMENTS

RESULT 1  
GTF1\_STRDO  
ID -GTF1\_STRDO STANDARD; PRT; 1597 AA.  
AC P11001;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 13-SEP-2005 (Rel. 48, Last annotation update)  
DE Glucosyltransferase-1 precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)  
DE (Sucrose 6-glucosyltransferase).  
GN Names=gtfi;  
OS Streptococcus downei (Streptococcus sobrinus).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=13117;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].  
RC STRAIN=MFE28; PubMed=3040686;  
RX MEDLINE=87308014; Gilpin M.L., Russell R.R.B.;  
RA Ferretti J.J., Gilpin M.L., Russell R.R.B.;  
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus  
sobrinus MFE28.";  
RL J. Bacteriol. 169:4271-4278(1987).  
CC -I- FUNCTION: Production of extracellular glucans, that are thought to  
play a key role in the development of the dental plaque because of  
their ability to adhere to smooth surfaces and mediate the  
aggregation of bacterial cells and food debris.  
CC -I- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-  
fructose + (1,6-alpha-D-glucosyl) (n+1).  
CC -I- SUBCELLULAR LOCATION: Secreted.  
CC -I- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha  
1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes  
water-soluble glucans (alpha 1,6-glucose). GTF-Si synthesizes both  
forms of glucans.  
CC -I- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.  
CC -I- SIMILARITY: Contains 19 cell wall binding repeats.

This Swiss-Prot entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use as long as its content is in no way modified and this statement is not  
removed.

EMBL; M17391; AAC63063.1; -; Genomic DNA.  
DR InterPro; IPR002479; Cell wall bd put.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW\_binding\_1; 4.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
DR Dental caries; Glucosyltransferase; Repeat; Signal; Transferase.  
FT SIGNAL 1 38 Potential.  
FT CHAIN 39 1597 Glucosyltransferase-I.  
FT REPEAT 1099 1132 A repeat.  
FT REPEAT 1163 1213 AC repeat.  
FT REPEAT 1227 1277 AC repeat.  
FT REPEAT 1292 1342 AC repeat.

```
FT REPEAT 1352 1399 B repeat.
FT REPEAT 1406 1455 AC repeat.
FT REPEAT 1465 1512 B repeat.
FT REPEAT 1519 1568 AC repeat.
FT REPEAT 1582 1597 A repeat (incomplete).
FT REGION 39 1050 Catalytic (approximate).
FT REGION 1099 1597 1.25 A, 2 B and 5 AC repeats.
FT REGION 1099 1597 Glucan-binding (approximate).
SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 100.0%; Score 113; DB 1; Length 1597;
Best Local Similarity 100.0%; Pred. No. 3.7e-09;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22
|||||:|||||:|||||:|||||:
Db 1303 TGAQTIKGQKLYFKANGQQVKG 1324

RESULT 2
Q59983_9STRE PRELIMINARY; PRT; 1590 AA.
AC Q59983;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5).
GN Name-gtfI;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OMZ176;
RA Sato S., Inoue M., Handa N., Aizawa Y., Iobse Y., Katayama T.;
RT "DNA sequence of the glucosyltransferase gene of serotype d
RT Streptococcus sobrinus.";
RL DNA Seq. 4:19-27(1993).
RN [2]
RP PROTEIN SEQUENCE.
RX MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
RT catalytic aspartic acid from two Streptococcus sobrinus alpha-
RT glucosyltransferases.";
RL J. Biol. Chem. 266:8916-8922(1991).
DR EMBL; D13858; BAA02976.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSSP; P06653; 1HCX.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_I; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Glycosyltransferase; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1590 glucosyltransferase-I.
SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 97.3%; Score 110; DB 2; Length 1590;
Best Local Similarity 95.5%; Pred. No. 1.2e-08;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22
|||||:|||||:|||||:|||||:
Db 1296 TGAQTIKGQKLYFKANGQQVKG 1317

RESULT 3
GTF2_STRDO STANDARD; PRT; 1592 AA.
ID GTF2_STRDO
AC P27470;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=6715 / Serotype G;
RX MEDLINE=91123227; PubMed=1704006;
RA Abo H., Matsumura T., Kodama T., Ohta H., Fukui K., Kato K.,
RA Kagawa H.;
RT "Peptide sequences for sucrose splitting and glucan binding within
RT Streptococcus sobrinus glucosyltransferase (water-insoluble glucan
RT synthetase).";
RL J. Bacteriol. 173:989-996(1991).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 16 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; D90213; BAA14241.1; -; Genomic_DNA.
DR HSSP; P06653; 1GVN.
DR InterPro; IPR002479; Cell wall_bd_put.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_I; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
DR Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38 Potential.
FT CHAIN 39 1592 Glucosyltransferase-I.
FT REPEAT 1093 1142 1.
FT REPEAT 1158 1207 2.
FT REPEAT 1222 1272 3.
FT REPEAT 1287 1337 4.
FT REPEAT 1402 1451 5.
FT REPEAT 1514 1563 6.
FT REPEAT 1577 1592 7 (incomplete).
FT REGION 39 1044 Catalytic (approximate).
FT REGION 1093 1592 7 X tandem repeats.
FT REGION 1093 1592 Glucan-binding (approximate).
FT CHAIN 39 1590 glucosyltransferase-I.
SQ SEQUENCE 1592 AA; 176168 MW; BC0A66D079351ECF CRC64;

Query Match 97.3%; Score 110; DB 1; Length 1592;
Best Local Similarity 95.5%; Pred. No. 1.2e-08;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22
|||||:|||||:|||||:|||||:
Db 1298 TGAQTIKGQKLYFKANGQQVKG 1319

RESULT 4
```

```

Q55263_9STRE
ID Q55263_9STRE PRELIMINARY; PRT; 1590 AA.
AC Q55263;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GTF-I.
GN Name=Glucosyltransferase;
OS Streptococcus sobrinus.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1310;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=ATCC 33478;
RA Sato S.;
RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase
produced from Streptococcus sobrinus ATCC 33478."
RL Kagoshima Daigaku Shigakubu Kiyo 16:23-29(1996).
RN [2]
RP PROTEIN SEQUENCE.
RC MEDLINE=91224988; PubMed=1827439;
RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;
RT "Isolation and sequence of an active-site peptide containing a
catalytic aspartic acid from two Streptococcus sobrinus alpha-
glucosyltransferases."
RL J. Biol. Chem. 266:8916-8922(1991).
DR ENBL; D63570; BAA09792.1; -; Genomic_DNA.
DR PIR; A39841; A39841.
DR HSP; P06653; LGVM.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3P2C6E4FD43 CRC64;

Query Match 86.7%; Score 98; DB 2; Length 1590;
Best Local Similarity 86.4%; Pred. No. 1.1e-06;
Matches 19; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGAQTIGQKLYFKANGQQVKG 22
:||||:||||:||||:||||:
Db 1296 SGAQTIGQKLYFKANGQQVKG 1317

RESULT 5
Q69A94_LEUME
ID Q69A94_LEUME PRELIMINARY; PRT; 1454 AA.
AC Q69A94;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Dextranucrase (EC 2.4.1.5).
GN Name=drp;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IBT-PQ;
RA Fernandez-Vazquez J.L., Lopez-Munigua A., Olvera C.;
RT "Molecular characterization of a dextranucrase gene from Leuconostoc
mesenteroides IBT-PQ isolated from pulque."
RL Submitted (DEC-2003) to the ENBL/GenBank/DBJ databases.
DR ENBL; AY504865; AAS79426.1; -; Genomic_DNA.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 1.
DR Pfam; PF02324; Glyco_hydro_70; 1.

```

```

KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1454 AA; 160569 MW; 2DC7342963B6FD87 CRC64;

Query Match 81.4%; Score 92; DB 2; Length 1454;
Best Local Similarity 77.3%; Pred. No. 9.7e-06;
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAQTIGQKLYFKANGQQVKG 22
||||| ||||: ||||: ||||: ||||:
Db 1280 TGAQTISGQKLFFDADGQIQK 1301

RESULT 6
Q8G9Q2_LEUME
ID Q8G9Q2_LEUME PRELIMINARY; PRT; 2835 AA.
AC Q8G9Q2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Dextranucrase (EC 2.4.1.5) (Fragment).
GN Name=dsrE;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC MEDLINE=22231661; PubMed=12270834;
RX DOI=10.1128/JB.184.20.5753-5761.2002;
RA Bozonnet S., Dols-Lafargue M., Fabre E., Pizzut S., Rемаud-Simeon M.,
Mondan P., Willemot R.M.;
RT "Molecular characterisation of DSR-E, an alpha-1,2 linkage
synthesising dextranucrase with two catalytic domains."
RL J. Bacteriol. 184:5753-5761(2002).
DR ENBL; AJ430204; CAD22883.1; -; Genomic DNA.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 3.
DR Pfam; PF02324; Glyco_hydro_70; 2.
KW Glycosyltransferase; Transferase.
FT NON_TER 1
SQ SEQUENCE 2835 AA; 313265 MW; D03262CDD735399D CRC64;

Query Match 81.4%; Score 92; DB 2; Length 2835;
Best Local Similarity 81.8%; Pred. No. 2e-05;
Matches 18; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAQTIGQKLYFKANGQQVKG 22
||||| ||||: ||||: ||||: ||||:
Db 1644 TGAQTINGQLYFKADGQQVKG 1665

RESULT 7
GTFB_STRMU
ID GTFB_STRMU STANDARD; PRT; 1476 AA.
AC P08987; O69381; O69384; O69387; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Suctrose 6-glucosyltransferase).
GN Name=gtfB; OrderedLocusNames=SMU.1004;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;

```

"Sequence analysis of the gtfB gene from Streptococcus mutans.";

RT J. Bacteriol. 169:4263-4270(1987).

RL [2]

RN NUCLEOTIDE SEQUENCE.

RP STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,

RC MT4467 / Serotype e, and MT8148 / Serotype c;

RD MEDLINE=98231643; PubMed=9570124;

RE Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,

RA Kimura S., Hamada S.;

RT "Molecular analyses of glucosyltransferase genes among strains of

RT Streptococcus mutans.";

RL FEMS Microbiol. Lett. 161:331-336(1998).

RN [3]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=UA159 / ATCC 700610 / Serotype C;

RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;

RA Ajdic D.J., McShan L.E., McLaughlin R.E., Savic G., Chang J.,

RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,

RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,

RA Perretti J.J.;

RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental

RT pathogen.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).

CC -1- FUNCTION: Production of extracellular glucans, that are thought to

CC play a key role in the development of the dental plaque because of

CC their ability to adhere to smooth surfaces and mediate the

CC aggregation of bacterial cells and food debris.

CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-

CC fructose + (1,6-alpha-D-glucosyl) (n+1).

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha

CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes

CC water-soluble glucans (alpha 1,6-glucose). GTF-S synthesizes both

CC forms of glucans.

CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.

CC -1- SIMILARITY: Contains 10 cell wall binding repeats.

CC -----

CC This Swiss-Prot entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use as long as its content is in no way modified and this statement is not

CC removed.

CC -----

DR EMBL; M17361; AAA88588.1; -; Genomic DNA.

DR EMBL; D88651; BAA26101.1; -; Genomic DNA.

DR EMBL; D88654; BAA26105.1; -; Genomic DNA.

DR EMBL; D88657; BAA26109.1; -; Genomic DNA.

DR EMBL; D88660; BAA26113.1; -; Genomic DNA.

DR EMBL; D89977; BAA26119.1; -; Genomic DNA.

DR EMBL; AE014940; AAN58705.1; -; Genomic DNA.

DR PIR; B33135; B33135.

DR HSP; P06653; IH8G.

DR InterPro; IPR002479; Cell\_wall\_bd\_put.

DR InterPro; IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding\_I; 4.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;

KW Transferase.

FT SIGNAL 1 34 Potential.

FT CHAIN 35 1476 Glucosyltransferase-I.

FT REPEAT 1097 1130 A repeat.

FT REPEAT 1161 1210 1.

FT REPEAT 1225 1275 2.

FT REPEAT 1290 1340 3.

FT REPEAT 1355 1405 4.

FT REPEAT 1420 1470 5.

FT REGION 35 1051 Catalytic (approximate).

FT REGION 1097 1476 Glucan-binding (approximate).

FT REGION 1161 1470 5 X tandem repeats.

FT VARIANT 62 62 S -> T (in strain MT4239).

FT VARIANT 65 65 T -> I (in strain GS-5).

FT VARIANT 68 68 V -> A (in strain GS-5, strain MT4245, strain MT4251, strain MT4467 and strain

MT8148).

FT VARIANT 78 78 Q -> P (in strain MT4251).

FT VARIANT 86 86 I -> S (in strain GS-5, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

FT VARIANT 89 89 S -> F (in strain MT4251).

FT VARIANT 168 168 K -> N (in strain MT4251).

FT VARIANT 276 276 S -> D (in strain GS-5, strain MT4467 and strain MT8148).

FT VARIANT 399 399 N -> R (in strain MT4239).

FT VARIANT 474 474 I -> T (in strain MT4239).

FT VARIANT 512 512 K -> R (in strain MT8148).

FT VARIANT 519 519 F -> Y (in strain MT8148).

FT VARIANT 701 701 T -> I (in strain MT8148).

FT VARIANT 708 708 T -> V (in strain MT8148).

FT VARIANT 938 938 F -> L (in strain MT8148).

FT VARIANT 952 957 FGKPEVE -> YGPEVA (in strain GS-5, strain MT4239 and strain MT4467).

FT VARIANT 963 964 SV -> NT (in strain GS-5, strain MT4239 and strain MT4467).

FT VARIANT 968 970 ADS -> VDG (in strain GS-5, strain MT4239 and strain MT4467).

FT VARIANT 1086 1086 A -> T (in strain MT4239).

FT VARIANT 1158 1158 S -> N (in strain MT4239).

FT VARIANT 1163 1163 H -> Y (in strain MT4251).

FT VARIANT 1168 1168 E -> K (in strain MT8148).

FT VARIANT 1182 1182 Y -> C (in strain MT8148).

FT VARIANT 1234 1234 A -> P (in strain MT4239).

FT VARIANT 1263 1263 R -> H (in strain GS-5 and strain MT4467).

FT VARIANT 1263 1263 R -> P (in strain MT8148).

FT VARIANT 1264 1264 Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).

FT VARIANT 1272 1272 S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).

FT VARIANT 1329 1329 H -> Y (in strain GS-5 and strain MT4467).

FT VARIANT 1394 1394 Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).

FT VARIANT 1402 1402 S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).

FT VARIANT 1459 1459 Y -> H (in strain MT4467).

FT CONFLICT 570 570 R -> A (in Ref. 1).

FT CONFLICT 800 817 ADQDVRAASTAPSTDGK -> LIKMFALRLARPHQOMA (in Ref. 1).

FT CONFLICT 1310 1310 H -> L (in Ref. 1).

SQ SEQUENCE 1476 AA; 165847 MW; 9C6E09F731B4CBCF CRC64;

Query Match 78.8%; Score 89; DB 1; Length 1476;

Best Local Similarity 77.3%; Pred. No. 3.1e-05;

Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAQTIKGKLYFKANGQVKG 22

Db 1236 TGSQTINGQHLYFRANGVQVKG 1257

RESULT 8

Q00599 STRSL

ID Q00599 STRSL PRELIMINARY; PRT; 1599 AA.

AC Q00599;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)

DE Glucosyltransferase S.

GN Name=gtfK;

OS Streptococcus salivarius.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1304;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=ATCC 25975;



RX MEDLINE=93381463; PubMed=83711114;  
RA Giffard P.M., Allen D.M., Milward C.P., Simpson C.L., Jacques N.A.;  
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and  
RT evolution of the gtf genes of oral streptococci.";  
RL J. Gen. Microbiol. 139:1511-1522(1993).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC 25975;  
RX MEDLINE=92148377; PubMed=1838391;  
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;  
RT "Molecular characterization of a cluster of at least tow  
RT glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";  
RL J. Gen. Microbiol. 137:2577-2593(1991).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC 25975;  
RA Jacques N.;  
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC 25975;  
RA Giffard P.M., O'Connor D.M., Milward C.P., Simpson C.L., Jacques N.A.;  
RT "Sequence of the gtfK gene of Streptococcus salivarius ATCC 25975 and  
RT the evolution of the gtf genes of oral streptococci.";  
RL Mol. Microbiol. 0:0-0(1992).  
DR EMBL; Z11873; CAA77898.1; -; Genomic\_DNA.  
DR PIR; S22737; S22737.  
DR HSP; P06653; I48G.  
DR GO; GO:0016740; P:transferase activity; IEA.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
DR InterPro; IPR002479; CW binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW binding\_1; 3.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
DR Transference.  
SQ SEQUENCE 1599 AA; 176480 MW; 24B77869E152B707 CRC64;  
Query Match 77.0%; Score 87; DB 2; Length 1599;  
Best Local Similarity 77.3%; Pred. No. 7.2e-05;  
Matches 17; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 TGAQTIGKQKLYFKANGQQVKG 22  
||||| ||||| ||||| |||||  
DB 1402 TGAQVINGQHLVFDANGRVKG 1423  
RESULT 9  
Q00600\_STRSL PRELIMINARY; PRT; 1518 AA.  
AC Q00600;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)  
DE Glucosyltransferase-I.  
GN Name=gtfJ;  
OS Streptococcus salivarius.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1304;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC 25975;  
RX MEDLINE=92148377; PubMed=1838391;  
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;  
RT "Molecular characterization of a cluster of at least tow  
RT glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";  
RL J. Gen. Microbiol. 137:2577-2593(1991).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=ATCC 25975;  
RA Jacques N.;  
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z11873; CAA77900.1; -; Genomic\_DNA.  
DR EMBL; M64111; AAA26896.1; -; Genomic\_DNA.  
DR PIR; A44811; A44811.  
DR HSP; P06653; IGVM.  
DR GO; GO:0016740; P:transferase activity; IEA.  
DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
DR InterPro; IPR002479; CW binding.  
DR InterPro; IPR003318; Glyco\_hydro\_70.  
DR Pfam; PF01473; CW binding\_1; 2.  
DR Pfam; PF02324; Glyco\_hydro\_70; 1.  
DR Transference.  
SQ SEQUENCE 1518 AA; 167730 MW; DAA41F717098B59A CRC64;  
Query Match 73.5%; Score 83; DB 2; Length 1518;  
Best Local Similarity 72.7%; Pred. No. 0.00031;  
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
QY 1 TGAQTIGKQKLYFKANGQQVKG 22  
||||| ||||| ||||| |||||  
DB 1318 TGAQVINGQKLYFNEDGSQVKG 1339  
RESULT 10  
GTFD\_STRMU STANDARD; PRT; 1462 AA.  
AC P49331; O69383; O69386; O69389; O69392; O69398;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-MAY-2005 (Rel. 47, Last annotation update)  
DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)  
DE (Sucrose 6-glucosyltransferase).  
GN Name=gtfd; OrderedLocustNames=SMU.910;  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=GS-5;  
RX MEDLINE=91100958; PubMed=2148600;  
RA Honda O., Kato C., Kuramitsu H.K.;  
RT "Nucleotide sequence of the Streptococcus mutans gtfD gene encoding  
RT the glucosyltransferase-S enzyme.";  
RL J. Gen. Microbiol. 136:2099-2105(1990).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,  
RC MT4467 / Serotype e, and MT8148 / Serotype C;  
RX MEDLINE=98231643; PubMed=9570124;  
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,  
RA Kimura S., Hamada S.;  
RT "Molecular analyses of glucosyltransferase genes among strains of  
RT Streptococcus mutans.";  
RL FEMS Microbiol. Lett. 161:331-336(1998).  
RN [3]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RC STRAIN=UA159 / ATCC 700610 / Serotype c;  
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;  
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,  
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,  
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lal H., White J., Roe B.A.,  
RA Ferretti J.J.;  
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental  
RT pathogen.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).  
CC -!- FUNCTION: Production of extracellular glucans, that are thought to  
CC play a key role in the development of the dental plaque because of  
CC their ability to adhere to smooth surfaces and mediate the  
CC aggregation of bacterial cells and food debris.  
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-  
CC fructose + (1,6-alpha-D-glucosyl)(n+1).  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha

1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both forms of glucans.

-/- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.

-/- SIMILARITY: Contains 6 cell wall binding repeats.

This Swiss-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use as long as its content is in no way modified and this statement is not removed.

ENBL; M29296; AAA26895.1; -; Genomic DNA.

ENBL; D86653; BAA26103.1; -; Genomic DNA.

ENBL; D86656; BAA26107.1; -; Genomic DNA.

ENBL; D86659; BAA26111.1; -; Genomic DNA.

ENBL; D86662; BAA26115.1; -; Genomic DNA.

ENBL; D89979; BAA26121.1; -; Genomic DNA.

ENBL; AEO14932; AAN58619.1; -; Genomic DNA.

HSSP; P06653; 1GVM.

InterPro; IPR002479; Cell wall bd put.

InterPro; IPR003318; Glyco\_hydro\_70.

Pfam; PF01473; CW\_binding\_1; 4.

Pfam; PF02324; Glyco\_hydro\_70; 1.

Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal; Transferrase.

SIGNAL 1 ? Potential.

CHAIN 1 1462 Glucosyltransferase-S.

REPEAT 1232 1295 1.

REPEAT 1296 1359 2.

REPEAT 1360 1423 3.

REGION 1232 1423

VARIANT 10 10

3 X 63 AA approximate tandem repeats.

Y -> H (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

I -> V (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

K -> E (in strain MT4467).

A -> S (in strain MT4239 and strain MT4245).

A -> T (in strain MT4251 and strain MT8148).

T -> I (in strain MT4239 and strain MT4245).

A -> V (in strain MT4239, strain MT4245, strain MT4251, strain MT4467 and strain MT8148).

V -> L (in strain MT8148).

E -> D (in strain MT8148).

D -> N (in strain MT4239, strain MT4245 and strain MT4251).

D -> N (in strain MT4239, strain MT4245 and strain MT4251).

Q -> H (in strain MT4245).

Q -> N (in strain MT4239 and strain MT4251).

E -> K (in strain MT4239).

V -> F (in strain MT4239).

F -> L (in strain MT4239, strain MT4251 and strain MT4467).

KKEYTO -> EKEYTL (in strain MT4251).

A -> S (in strain MT4239).

TDGSEA -> ADKGND (in strain MT4251).

TDGS -> ADKGN (in strain MT4239 and strain MT4245).

T -> A (in strain GS-5, strain MT4239, strain MT4245, strain MT4251, strain

FT VARIANT 964 964 MT4467 and strain MT8148).

FT VARIANT 1019 1019 D -> Y (in strain MT4251).

FT VARIANT 1019 1019 E -> K (in strain MT4245 and strain MT4251).

FT VARIANT 1059 1060 LG -> IR (in strain MT4251).

FT VARIANT 1060 1060 G -> R (in strain MT4245).

FT VARIANT 1080 1080 G -> R (in strain MT4239).

FT VARIANT 1142 1142 H -> Q (in strain GS-5).

FT VARIANT 1198 1198 S -> N (in strain MT4239).

FT VARIANT 1220 1220 Y -> C (in strain MT4251 and strain MT4467).

FT VARIANT 1280 1280 F -> L (in strain MT4467).

FT VARIANT 1282 1282 Q -> P (in strain MT4245).

FT VARIANT 1290 1290 K -> T (in strain MT4245).

FT VARIANT 1311 1311 N -> D (in strain MT4245).

FT VARIANT 1403 1403 G -> D (in strain GS-5 and strain MT4467).

FT VARIANT 1425 1425 G -> R (in strain GS-5).

FT VARIANT 1449 1449 R -> K (in strain MT4467).

FT CONFLICT 1428 1462 RYDKNSGNMVKVVTLANGRRIGIDRWGIARYY -> VY R (in Ref. 1).

SQ SEQUENCE 1462 AA; 163388 MW; CE4A279C4D708645 CRC64;

Query Match 72.6%; Score 82; DB 1; Length 1462;

Best Local Similarity 68.2%; Pred. NO. 0.0043;

Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGAQTIKGKLYFKANGQQVK 22

DB 1332 TGSQTIAAGKLYPASDGQVK 1353

RESULT 11

Q8VUH3\_STRMU PRELIMINARY; PRT; 591 AA.

ID Q8VUH3\_STRMU PRELIMINARY; PRT; 591 AA.

AC Q8VUH3;

DT 01-MAR-2002 (TrEMBLrel. 20, Created)

DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Glucosyltransferase SI (Fragment).

GN Name=gtfc;

OS Streptococcus mutans.

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1309;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=Xc;

RX MEDLINE=22322694; PubMed=12435673;

RX DOI=10.1128/AAC.46.12.3756-3764.2002;

RA Teuda H., Yamashita Y., Shibata Y., Nakano Y., Koga T.i

RT "Genes involved in bacitracin resistance in Streptococcus mutans."

RL Antimicrob. Agents Chemother. 46:3756-3764(2002).

DR EMBL; AB078507; BAB83942.1; -; Genomic DNA.

DR HSSP; P07762; 1M7X.

DR GO; GO:0016740; F:transferase activity; IEA.

DR GO; GO:0009250; P:glucan biosynthesis; IEA.

DR InterPro; IPR002479; CW\_binding.

DR InterPro; IPR003318; Glyco\_hydro\_70.

DR Pfam; PF01473; CW\_binding\_1; 2.

DR Pfam; PF02324; Glyco\_hydro\_70; 1.

KW Transferase.

FT NON\_TER 1 1

SQ SEQUENCE 591 AA; 67094 MW; 0933DCB4421DAF30 CRC64;

Query Match 71.7%; Score 81; DB 2; Length 591;

Best Local Similarity 63.6%; Pred. NO. 0.00024;

Matches 14; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAQTIKGKLYFKANGQQVK 22

DB 465 TGAUVNGQRLYFKSNGVQAKG 486

```

RESULT 12
Q9WXA4_9STRE
ID Q9WXA4_9STRE PRELIMINARY; PRT; 1338 AA.
AC Q9WXA4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GTF-S.
GN Name=gfts;
OS Streptococcus criceti.
OG Plasmid pAM1.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77236.1; -; Genomic_DNA.
DR HSSP; P06653; IH8G.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1338 AA; 148558 MW; 0A90C8E10E15D99B CRC64;

Query Match 71.7%; Score 81; DB 2; Length 1338;
Best Local Similarity 68.2%; Pred. No. 0.00059;
Matches 15; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAQTIGKQKLYFKANGQVKG 22
Db 1207 TGSQTIGAKQKVFQPNQVQVKG 1228

RESULT 13
Q9WXA5_9STRE
ID Q9WXA5_9STRE PRELIMINARY; PRT; 1512 AA.
AC Q9WXA5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GTF-S.
GN Name=gfts;
OS Streptococcus criceti.
OG Plasmid pAM1.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77237.1; -; Genomic_DNA.
DR HSSP; P06654; IMPE.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9CB601FC14 CRC64;

Query Match 70.4%; Score 79.5; DB 2; Length 1512;
Best Local Similarity 81.8%; Pred. No. 0.0012;
Matches 18; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 2 GAQTIGKQKLYFKANGQVKG 22

```

```

Db 1269 GAQTIGKQKLYFKANGQVKG 1290

RESULT 14
Q48756_LEUME
ID Q48756_LEUME PRELIMINARY; PRT; 1290 AA.
AC Q48756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Dextranucrase.
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NRRL B1299;
RL MEDLINE=97136686; PubMed=8982063; DOI=10.1016/S0378-1119(96)00443-X;
RA Monchois V., Willemot R.M., Renaud-Simeon M., Croux C., Monsan P.;
RT "Cloning and sequencing of a gene coding for a novel dextranucrase
from Leuconostoc mesenteroides NRRL B-1299 synthesizing only alpha (1-
6) and alpha (1-3) linkages.";
RL Gene 182:23-32(1996).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NRRL B1299;
RA Monchois V.M., Willemot R.M., Renaud-Simeon M., Croux C., Monsan P.;
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U38181; AAB40875.1; -; Genomic_DNA.
DR PIR; JCS473; JCS473.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 1.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1290 AA; 145590 MW; 3555C2E968749FAA CRC64;

Query Match 69.0%; Score 78; DB 2; Length 1290;
Best Local Similarity 68.2%; Pred. No. 0.0017;
Matches 15; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TGAQTIGKQKLYFKANGQVKG 22
Db 1245 TGVQVINGQTLTFDADGRQVKG 1266

RESULT 15
GTF5_STRDO
ID GTF5_STRDO STANDARD; PRT; 1365 AA.
AC P29336;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE Glucosyltransferase-S precursor (EC 2.4.1.5) (GTF-S) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gfts;
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=MPE28;
RL MEDLINE=90316665; PubMed=2142479;
RA Gilmore K.S., Russell R.R., Ferretti J.J.;
RT "Analysis of the Streptococcus downei gtfS gene, which specifies a
glucosyltransferase that synthesizes soluble glucans.";
RL Infect. Immun. 58:2452-2458(1990).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
play a key role in the development of the dental plaque because of
their ability to adhere to smooth surfaces and mediate the
aggregation of bacterial cells and food debris.
CC

```



Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	113	100.0	22	1	US-08-057-1628-3	Sequence 3, Appli
2	89	78.8	1475	2	US-09-007-999-2	Sequence 2, Appli
3	89	78.8	1475	2	US-09-210-361-2	Sequence 2, Appli
4	89	78.8	1475	2	US-09-740-274-2	Sequence 2, Appli
5	87	77.0	22	2	US-09-290-049A-3	Sequence 3, Appli
6	82	72.6	1430	2	US-09-008-172-2	Sequence 2, Appli
7	82	72.6	1430	2	US-09-210-361-6	Sequence 6, Appli
8	82	72.6	1430	2	US-09-740-274-6	Sequence 6, Appli
9	74	65.5	1375	2	US-09-210-361-4	Sequence 4, Appli
10	74	65.5	1375	2	US-09-740-274-4	Sequence 4, Appli
11	71	62.8	1577	1	US-08-793-824-2	Sequence 2, Appli
12	61	54.0	2057	2	US-09-499-203-2	Sequence 2, Appli
13	53	46.9	624	2	US-09-877-730-24	Sequence 24, Appl
14	53	46.9	712	2	US-09-877-730-22	Sequence 22, Appl
15	53	46.9	793	2	US-09-877-730-28	Sequence 28, Appl
16	53	46.9	991	2	US-09-877-730-12	Sequence 12, Appl
17	53	46.9	1069	2	US-09-877-730-2	Sequence 2, Appli
18	53	46.9	1072	2	US-09-877-730-18	Sequence 18, Appl
19	53	46.9	1150	2	US-09-877-730-8	Sequence 8, Appli
20	52	46.0	811	1	US-08-480-604A-7	Sequence 7, Appli
21	52	46.0	811	1	US-08-405-496A-7	Sequence 7, Appli
22	52	46.0	811	2	US-08-915-136-7	Sequence 7, Appli
23	52	46.0	811	2	US-08-957-310-7	Sequence 7, Appli
24	52	46.0	811	2	US-10-011-366-7	Sequence 7, Appli
25	52	46.0	811	2	US-09-084-517-7	Sequence 7, Appli
26	52	46.0	812	1	US-08-480-604A-29	Sequence 29, Appl
27	52	46.0	812	2	US-08-915-136-29	Sequence 29, Appl

```
Db 1 TGAQTIKGQKLYFKANGQQVKG 22

RESULT 2
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; PRIOR FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2

Query Match 78.8%; Score 89; DB 2; Length 1475;
Best Local Similarity 77.3%; Pred. No. 7e-06;
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22
||:|||||:|||||:|||||
Db 1235 TGSQTINGQHLYFRANGVQVKG 1256

RESULT 3
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; PRIOR FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2

Query Match 78.8%; Score 89; DB 2; Length 1475;
Best Local Similarity 77.3%; Pred. No. 7e-06;
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22
||:|||||:|||||:|||||
Db 1235 TGSQTINGQHLYFRANGVQVKG 1256

RESULT 4
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 78.8%; Score 89; DB 2; Length 1475;
Best Local Similarity 77.3%; Pred. No. 7e-06;
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22
||:|||||:|||||:|||||
Db 1235 TGSQTINGQHLYFRANGVQVKG 1256

RESULT 5
US-09-290-049A-3
; Sequence 3, Application US/09290049A
; Patent No. 6827936
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINES FOR DENTAL
; FILE REFERENCE: CRIES
; CURRENT APPLICATION NUMBER: US/09/290,049A
; CURRENT FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: GLB peptide
US-09-290-049A-3

Query Match 77.0%; Score 87; DB 2; Length 22;
Best Local Similarity 77.3%; Pred. No. 1.2e-07;
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22
||:|||||:|||||:|||||
```

```
Db      1  TGAQTNGQLLYFRANGVQVKG 22

RESULT 6
US-09-008-172-2
; Sequence 2, Application US/09008172
; Patent No. 6127602
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; FILE REFERENCE: 0358D
; CURRENT APPLICATION NUMBER: US/09/008,172
; PRIOR FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-008-172-2

Query Match      72.6%; Score 82; DB 2; Length 1430;
Best Local Similarity 68.2%; Pred. No. 9.7e-05;
Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy      1  TGAQTIGOKLYFKANGQOVKG 22
||:|||||:|||||:|||||:|||||
Db      1332 TGSQTIGAKKLYFASDGKQVKG 1353

RESULT 7
US-09-210-361-6
; Sequence 6, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; PRIOR FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-6

Query Match      72.6%; Score 82; DB 2; Length 1430;
Best Local Similarity 68.2%; Pred. No. 9.7e-05;
Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy      1  TGAQTIGOKLYFKANGQOVKG 22
||:|||||:|||||:|||||:|||||
Db      1332 TGSQTIGAKKLYFASDGKQVKG 1353

RESULT 8
US-09-740-274-6
; Sequence 6, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 1430
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-740-274-6

Query Match      72.6%; Score 82; DB 2; Length 1430;
Best Local Similarity 68.2%; Pred. No. 9.7e-05;
Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy      1  TGAQTIGOKLYFKANGQOVKG 22
||:|||||:|||||:|||||:|||||
Db      1332 TGSQTIGAKKLYFASDGKQVKG 1353

RESULT 9
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4

Query Match      65.5%; Score 74; DB 2; Length 1375;
```

Best Local Similarity 63.6%; Pred. No. 0.0019;  
Matches 14; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22  
Db 1264 TGTVTFGQRLYFKPVGQAKG 1285

RESULT 10  
US-09-740-274-4  
; Sequence 4, Application US/09740274  
; Patent No. 6465203  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 65.5%; Score 74; DB 2; Length 1375;  
Best Local Similarity 63.6%; Pred. No. 0.0019;  
Matches 14; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22  
Db 1264 TGTVTFGQRLYFKPVGQAKG 1285

RESULT 11  
US-08-793-824-2  
; Sequence 2, Application US/08793824  
; Patent No. 5981838  
; GENERAL INFORMATION:  
; APPLICANT: Simpson, Christine Lynn  
; APPLICANT: Giffard, Philip Morrison  
; APPLICANT: Jacques, Nicholas Anthony  
; TITLE OF INVENTION: Genetic Manipulation of Plants to  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Griffith Hack & Co  
; STREET: Level 8, 168 Walker Street  
; CITY: No. 5981838th Sydney  
; STATE: New South Wales  
; COUNTRY: Australia  
; ZIP: 2060  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/793,824  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: AU PM7643  
; FILING DATE: 24-AUG-1994  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 61 2 9957 5944  
; TELEFAX: 61 2 957 6288  
; TELEX: 26547  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1577 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Streptococcus salivarius  
US-08-793-824-2

Query Match 62.8%; Score 71; DB 1; Length 1577;  
Best Local Similarity 59.1%; Pred. No. 0.0071;  
Matches 13; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22  
Db 1533 TGOOVINGKQLYFDGSGRQVKG 1554

RESULT 12  
US-09-499-203-2  
; Sequence 2, Application US/09499203  
; Patent No. 6570065  
; GENERAL INFORMATION:  
; APPLICANT: KOSSMANN, Jens  
; APPLICANT: WELSH, Thomas  
; APPLICANT: QUANZ, Martin  
; APPLICANT: KNUTH, Karola  
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase  
; FILE REFERENCE: 147-196P  
; CURRENT APPLICATION NUMBER: US/09/499,203  
; CURRENT FILING DATE: 2000-02-08  
; NUMBER OF SEQ ID NOS: 54  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 2057  
; TYPE: PRT  
; ORGANISM: Leuconostoc mesenteroides  
US-09-499-203-2

Query Match 54.0%; Score 61; DB 2; Length 2057;  
Best Local Similarity 59.1%; Pred. No. 0.43;  
Matches 13; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22  
Db 176 TGLQNIIDGNLQYFDDNGYQVKG 197

RESULT 13  
US-09-877-730-24  
; Sequence 24, Application US/09877730  
; Patent No. 6465632  
; GENERAL INFORMATION:  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Scoville, John  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the



```
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 624
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-877-730-24
```

```
Query Match      46.9%; Score 53; DB 2; Length 624;
Best Local Similarity 58.8%; Pred. No. 2.3;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

```
QY      6 IKGQKLYFKANGQOVKG 22
      |:|:|:|:|:|:|
Db      203 IQGYKLYYKEEGQQENG 219
```

## RESULT 14

```
US-09-877-730-22
; Sequence 22, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 712
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-877-730-22
```

```
Query Match      46.9%; Score 53; DB 2; Length 712;
Best Local Similarity 58.8%; Pred. No. 2.7;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

```
QY      6 IKGQKLYFKANGQOVKG 22
      |:|:|:|:|:|:|
Db      291 IQGYKLYYKEEGQQENG 307
```

## RESULT 15

```
US-09-877-730-28
; Sequence 28, Application US/09877730
; Patent No. 6465632
; GENERAL INFORMATION:
; APPLICANT: Walke, D. Wade
; APPLICANT: Scoville, John
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. 6465632el Human Phosphatases and Polynucleotides Encoding the
; FILE REFERENCE: LEX-0189-USA
; CURRENT APPLICATION NUMBER: US/09/877,730
; CURRENT FILING DATE: 2001-06-08
```

```
; PRIOR APPLICATION NUMBER: US 60/210,607
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 28
; LENGTH: 793
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-877-730-28
```

```
Query Match      46.9%; Score 53; DB 2; Length 793;
Best Local Similarity 58.8%; Pred. No. 3;
Matches 10; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
```

```
QY      6 IKGQKLYFKANGQOVKG 22
      |:|:|:|:|:|:|
Db      291 IQGYKLYYKEEGQQENG 307
```

```
Search completed: February 10, 2006, 22:36:13
Job time : 4.96976 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:14:04 ; Search time 12.9073 Seconds  
(without alignments)  
712.176 Million cell updates/sec

Title: US-10-797-821-23

Perfect score: 113

Sequence: 1 TGAQTIGKGLYFKANGQQVKG 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA Main:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	113	100.0	22	4	US-10-383-930-23
2	113	100.0	22	5	US-10-797-821-23
3	110	97.3	1590	4	US-10-383-930-37
4	110	97.3	1590	5	US-10-797-821-37
5	89	78.8	1475	3	US-09-740-274-2
6	89	78.8	1475	4	US-10-383-930-34
7	89	78.8	1475	5	US-10-797-821-34
8	87	77.0	22	5	US-10-797-302-3
9	83	73.5	1518	4	US-10-383-930-40
10	83	73.5	1518	5	US-10-797-821-40
11	82	72.6	1430	3	US-09-740-274-6
12	82	72.6	1430	4	US-10-383-930-36
13	82	72.6	1430	5	US-10-797-821-36
14	78	69.0	1365	4	US-10-383-930-39
15	78	69.0	1365	5	US-10-797-821-39
16	74	65.5	1375	3	US-09-740-274-4
17	74	65.5	1375	4	US-10-383-930-35
18	74	65.5	1375	5	US-10-797-821-35
19	73	64.6	726	4	US-10-282-122A-72033
20	66	58.4	1497	5	US-10-484-218-18
21	65	57.5	1066	5	US-10-484-218-22
22	61	54.0	2057	4	US-10-417-280A-2
23	58	51.3	1595	5	US-10-484-218-20
24	53	46.9	624	4	US-10-231-353-24
25	53	46.9	624	6	US-11-128-073-24
26	53	46.9	673	5	US-10-496-905-516
27	53	46.9	673	5	US-10-496-905-519

28	53	46.9	712	4	US-10-231-353-22	Sequence 22, Appl
29	53	46.9	712	6	US-11-128-073-22	Sequence 22, Appl
30	53	46.9	793	4	US-10-231-353-28	Sequence 28, Appl
31	53	46.9	793	6	US-11-128-073-28	Sequence 28, Appl
32	53	46.9	991	4	US-10-231-353-12	Sequence 12, Appl
33	53	46.9	991	6	US-11-128-073-12	Sequence 12, Appl
34	53	46.9	1069	4	US-10-231-353-2	Sequence 2, Appl
35	53	46.9	1069	6	US-11-128-073-2	Sequence 2, Appl
36	53	46.9	1072	4	US-10-231-353-18	Sequence 18, Appl
37	53	46.9	1072	6	US-11-128-073-18	Sequence 18, Appl
38	53	46.9	1109	5	US-10-450-763-42830	Sequence 42830, A
39	53	46.9	1150	4	US-10-231-353-8	Sequence 8, Appl
40	53	46.9	1150	6	US-11-128-073-8	Sequence 8, Appl
41	52.5	46.5	1554	4	US-10-383-930-38	Sequence 38, Appl
42	52.5	46.5	1554	5	US-10-797-821-38	Sequence 38, Appl
43	52	46.0	811	4	US-10-011-366-7	Sequence 7, Appl
44	52	46.0	811	4	US-10-354-774-7	Sequence 7, Appl
45	52	46.0	811	4	US-10-271-012-7	Sequence 7, Appl

ALIGNMENTS

RESULT 1  
US-10-383-930-23  
; Sequence 23, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patent in version 3.2  
; SEQ ID NO 23  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-23

Query Match 100.0%; Score 113; DB 4; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1.7e-10;  
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGAQTIGKGLYFKANGQQVKG 22  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 TGAQTIGKGLYFKANGQQVKG 22

RESULT 2  
US-10-797-821-23  
; Sequence 23, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049

```
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 23
; LENGTH: 22
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: GTF-derived glucan binding (GLU) peptide
US-10-797-821-23

Query Match          100.0%; Score 113; DB 5; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.7e-10;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22
   |||||:|||||:|||||:|||||
Db 1 TGAQTIRGQKLYFKANGQQVKG 22

RESULT 3
US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match          97.3%; Score 110; DB 4; Length 1590;
Best Local Similarity 95.5%; Pred. No. 5.3e-08;
Matches 21; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22
   |||||:|||||:|||||:|||||
Db 1296 TGAQTIRGQKLYFKANGQQVKG 1317

RESULT 4
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; FILE REFERENCE: 25669-018
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/210,361
; CURRENT FILING DATE: 2000-12-19
; CURRENT APPLICATION NUMBER: US/09/740,274
; FILE REFERENCE: 0357CRD
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; APPLICANT: Nichols, Scott E.
; GENERAL INFORMATION:
; Patent No. US20020031826A1
; Sequence 2, Application US/09740274
; Patent No. US20020031826A1
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match          78.8%; Score 89; DB 3; Length 1475;
Best Local Similarity 77.3%; Pred. No. 9.4e-05;
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAQTIKGQKLYFKANGQQVKG 22
   ||:|||||:|||||:|||||
Db 1235 TGSQTINGQHLYFRANGVQVKG 1256

RESULT 6
US-10-383-930-34
; Sequence 34, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
```

; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 34  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-34

Query Match 78.8%; Score 89; DB 4; Length 1475;  
Best Local Similarity 77.3%; Pred. No. 9.4e-05;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAQTIGOKLYFKANGQOVKG 22  
||:|||||:|||||  
Db 1235 TGSQTINGQHLFYFRANGVQVKG 1256

RESULT 7  
US-10-797-821-34  
; Sequence 34, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 34  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-34

Query Match 78.8%; Score 89; DB 5; Length 1475;  
Best Local Similarity 77.3%; Pred. No. 9.4e-05;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAQTIGOKLYFKANGQOVKG 22  
||:|||||:|||||  
Db 1235 TGSQTINGQHLFYFRANGVQVKG 1256

RESULT 8  
US-10-797-302-3  
; Sequence 3, Application US/10797302  
; Publication No. US20050026271A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-019  
; CURRENT APPLICATION NUMBER: US/10/797,302

; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-02  
; NUMBER OF SEQ ID NOS: 21  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 3  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: GLB peptide  
US-10-797-302-3

Query Match 77.0%; Score 87; DB 5; Length 22;  
Best Local Similarity 77.3%; Pred. No. 1.9e-06;  
Matches 17; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGAQTIGOKLYFKANGQOVKG 22  
||:|||||:|||||  
Db 1 TGAQTINGQLLYFRANGVQVKG 22

RESULT 9  
US-10-383-930-40  
; Sequence 40, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 40  
; LENGTH: 1518  
; TYPE: PRT  
; ORGANISM: Streptococcus salivarius  
US-10-383-930-40

Query Match 73.5%; Score 83; DB 4; Length 1518;  
Best Local Similarity 72.7%; Pred. No. 0.00084;  
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TGAQTIGOKLYFKANGQOVKG 22  
||:|||||:|||||  
Db 1318 TGAQVINGQKLYFNEDGSQVKG 1339

RESULT 10  
US-10-797-821-40  
; Sequence 40, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020

US-10-383-930-36  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 40  
; LENGTH: 1518  
; TYPE: PR1  
; ORGANISM: Streptococcus salivarius  
US-10-797-821-40

Query Match 73.5%; Score 83; DB 5; Length 1518;  
Best Local Similarity 72.7%; Pred. No. 0.00084;  
Matches 16; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TGAQTIGOKLYFKANGQOVKG 22  
|||:|||||:|||||:|||||  
Db 1318 TGAQVINGOKLYFNEDGSQVKG 1339

RESULT 11  
US-09-740-274-6  
; Sequence 6, Application US/09740274  
; Patent No. US20020031826A1  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 6  
; LENGTH: 1430  
; TYPE: PR1  
; ORGANISM: streptococcus mutans  
US-09-740-274-6

Query Match 72.6%; Score 82; DB 3; Length 1430;  
Best Local Similarity 68.2%; Pred. No. 0.0011;  
Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAQTIGOKLYFKANGQOVKG 22  
|||:|||||:|||||:|||||  
Db 1332 TGSQTIAGKKLYFASDGQVKG 1353

RESULT 12

US-10-383-930-36  
; Sequence 36, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 36  
; LENGTH: 1430  
; TYPE: PR1  
; ORGANISM: Streptococcus mutans  
US-10-383-930-36

Query Match 72.6%; Score 82; DB 4; Length 1430;  
Best Local Similarity 68.2%; Pred. No. 0.0011;  
Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAQTIGOKLYFKANGQOVKG 22  
|||:|||||:|||||:|||||  
Db 1332 TGSQTIAGKKLYFASDGQVKG 1353

RESULT 13  
US-10-797-821-36  
; Sequence 36, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 36  
; LENGTH: 1430  
; TYPE: PR1  
; ORGANISM: Streptococcus mutans  
US-10-797-821-36

Query Match 72.6%; Score 82; DB 5; Length 1430;  
Best Local Similarity 68.2%; Pred. No. 0.0011;  
Matches 15; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGAQTIGOKLYFKANGQOVKG 22  
|||:|||||:|||||:|||||  
Db 1332 TGSQTIAGKKLYFASDGQVKG 1353

RESULT 14  
US-10-383-930-39

; Sequence 39, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 39  
; LENGTH: 1365  
; TYPE: PRT  
; ORGANISM: Streptococcus downei  
US-10-383-930-39

Query Match 69.0%; Score 78; DB 4; Length 1365;  
Best Local Similarity 68.2%; Pred. No. 0.0045;  
Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAQTIGQKLYFKANGQQVKG 22  
||| ||| : : |||  
Db 1235 TGEQTIDGQKVFQDNGVQVKG 1256

RESULT 15  
US-10-797-821-39  
; Sequence 39, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 39  
; LENGTH: 1365  
; TYPE: PRT  
; ORGANISM: Streptococcus downei  
US-10-797-821-39

Query Match 69.0%; Score 78; DB 5; Length 1365;  
Best Local Similarity 68.2%; Pred. No. 0.0045;  
Matches 15; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 TGAQTIGQKLYFKANGQQVKG 22  
||| ||| : : |||  
Db 1235 TGEQTIDGQKVFQDNGVQVKG 1256

Search completed: February 10, 2006, 23:25:09  
Job time : 13.9073 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:15:51 ; Search time 0.953629 Seconds  
(without alignments)  
302.734 Million cell updates/sec

Title: US-10-797-821-23

Perfect score: 113

Sequence: 1 TGAQTIKGKLYPKANGQQVKG 22

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database :
- 1: /cgn2\_6/prodata/1/pubppa/US08\_NEW\_PUB pep.\*
  - 2: /cgn2\_6/prodata/1/pubppa/US06\_NEW\_PUB pep.\*
  - 3: /cgn2\_6/prodata/1/pubppa/US07\_NEW\_PUB pep.\*
  - 4: /cgn2\_6/prodata/1/pubppa/PCT\_NEW\_PUB pep.\*
  - 5: /cgn2\_6/prodata/1/pubppa/US09\_NEW\_PUB pep.\*
  - 6: /cgn2\_6/prodata/1/pubppa/US10\_NEW\_PUB pep.\*
  - 7: /cgn2\_6/prodata/1/pubppa/US11\_NEW\_PUB pep.\*
  - 8: /cgn2\_6/prodata/1/pubppa/US60\_NEW\_PUB pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	73	64.6	726	7	US-11-052-554A-217
2	53	46.9	1150	7	US-11-139-435-1
3	52	46.0	2710	7	US-11-051-453-41
4	51	45.1	396	7	US-11-022-562-228
5	49	43.4	1191	7	US-11-139-435-2
6	49	43.4	1193	7	US-11-139-435-3
7	47	41.6	710	6	US-10-467-657-4292
8	44	38.9	592	7	US-11-052-554A-235
9	44	38.9	592	7	US-11-212-443-91
10	44	38.9	592	7	US-11-212-443-98
11	44	38.9	592	7	US-11-212-443-106
12	44	38.9	592	7	US-11-212-443-107
13	44	38.9	592	7	US-11-212-443-110
14	44	38.9	600	7	US-11-212-443-111
15	44	38.9	1042	7	US-11-067-811-1
16	42	37.2	291	7	US-11-212-443-138
17	42	37.2	502	6	US-10-131-826A-548
18	42	37.2	502	6	US-10-689-742-148
19	42	37.2	592	7	US-11-212-443-6
20	41	36.3	1275	6	US-10-821-234-1598
21	40	35.4	211	6	US-10-467-657-6932
22	40	35.4	594	7	US-11-212-443-93
23	40	35.4	594	7	US-11-212-443-99
24	40	35.4	594	7	US-11-212-443-102
25	39	34.5	202	6	US-10-793-626-2364

ALIGNMENTS

RESULT 1  
US-11-052-554A-217  
; Sequence 217, Application US/11052554A  
; Publication No. US2005028866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR APPLICATION NUMBER: US 60/589,227  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 217  
; LENGTH: 726  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans UA159  
US-11-052-554A-217

Query Match 64.6%; Score 73; DB 7; Length 726;  
Best Local Similarity 59.1%; Pred. No. 0.00053;  
Matches 13; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

Qy 1 TGAQTIKGKLYPKANGQQVKG 22  
Db 235 TGSITVNGOSLYFNSDGSQVKG 256

RESULT 2  
US-11-139-435-1  
; Sequence 1, Application US/11139435  
; Publication No. US2005028766A1  
; GENERAL INFORMATION:  
; APPLICANT: Fann, Ming-Ji  
; TITLE OF INVENTION: Cellular Proliferation Control Factors  
; FILE REFERENCE: 17741-002001  
; CURRENT APPLICATION NUMBER: US/11/139,435  
; CURRENT FILING DATE: 2005-05-27  
; PRIOR APPLICATION NUMBER: US 60/575,611  
; PRIOR FILING DATE: 2004-05-27  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 1150

Sequence 230, App  
Sequence 92, Appl  
Sequence 94, Appl  
Sequence 101, App  
Sequence 100, App  
Sequence 17, Appl  
Sequence 155, App  
Sequence 184, App  
Sequence 368, App  
Sequence 6470, Ap  
Sequence 2, Appli  
Sequence 190, App  
Sequence 5980, Ap  
Sequence 4, Appli  
Sequence 95, Appl  
Sequence 96, Appl  
Sequence 103, App  
Sequence 104, App  
Sequence 105, App  
Sequence 229, App



```
US-10-467-657-4292
; Sequence 4292, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 4292
; LENGTH: 710
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-4292

Query Match      41.6%; Score 47; DB 6; Length 710;
Best Local Similarity 44.4%; Pred. No. 6.8;
Matches 8; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

QY      2 GAOTIKGQKLYFKANGQ 19
Db      76 QSTEGEEIYFRNNSQ 93

RESULT 8
US-11-052-554A-235
; Sequence 235, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 235
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Neisseria meningitidis Z2491
US-11-052-554A-235

Query Match      38.9%; Score 44; DB 7; Length 592;
Best Local Similarity 45.0%; Pred. No. 17;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY      1 TGAQTIKGQKLYFKANGQ 20
Db      134 TGLINVETEKLSFGANGKV 153

RESULT 9
US-11-212-443-91
; Sequence 91, Application US/11212443
; Publication No. US20050287165A1
; GENERAL INFORMATION:
; APPLICANT: Scalato, Enzo
; APPLICANT: Masignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: PIZZA, Mariagrazia
; TITLE OF INVENTION: Meningococcal Antigens
; FILE REFERENCE: CHIR0159
```

```
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Meningococcal Antigens
; FILE REFERENCE: CHIR0159
; CURRENT APPLICATION NUMBER: US/11/212,443
; CURRENT FILING DATE: 2005-08-24
; PRIOR APPLICATION NUMBER: US/09/302,626
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/IB99/00103
; PRIOR FILING DATE: 1999-01-14
; NUMBER OF SEQ ID NOS: 195
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 91
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-11-212-443-91

Query Match      38.9%; Score 44; DB 7; Length 592;
Best Local Similarity 45.0%; Pred. No. 17;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY      1 TGAQTIKGQKLYFKANGQ 20
Db      134 TGLINVETEKLSFGANGKV 153

RESULT 10
US-11-212-443-98
; Sequence 98, Application US/11212443
; Publication No. US20050287165A1
; GENERAL INFORMATION:
; APPLICANT: Scalato, Enzo
; APPLICANT: Masignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: PIZZA, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Meningococcal Antigens
; FILE REFERENCE: CHIR0159
; CURRENT APPLICATION NUMBER: US/11/212,443
; CURRENT FILING DATE: 2005-08-24
; PRIOR APPLICATION NUMBER: US/09/302,626
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/IB99/00103
; PRIOR FILING DATE: 1999-01-14
; NUMBER OF SEQ ID NOS: 195
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 98
; LENGTH: 592
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-11-212-443-98

Query Match      38.9%; Score 44; DB 7; Length 592;
Best Local Similarity 45.0%; Pred. No. 17;
Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY      1 TGAQTIKGQKLYFKANGQ 20
Db      134 TGLINVETEKLSFGANGKV 153

RESULT 11
US-11-212-443-106
; Sequence 106, Application US/11212443
; Publication No. US20050287165A1
; GENERAL INFORMATION:
; APPLICANT: Scalato, Enzo
; APPLICANT: Masignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: PIZZA, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Meningococcal Antigens
; FILE REFERENCE: CHIR0159
```

```
; CURRENT APPLICATION NUMBER: US/11/212,443
; CURRENT FILING DATE: 2005-08-24
; PRIOR APPLICATION NUMBER: US/09/302,626
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/IB99/00103
; PRIOR FILING DATE: 1999-01-14
; NUMBER OF SEQ ID NOS: 195
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 106
; TYPE: PRT
; LENGTH: 592
; ORGANISM: Neisseria meningitidis
US-11-212-443-106

Query Match      38.9%; Score 44; DB 7; Length 592;
Best Local Similarity 45.0%; Pred. No. 17;
Matches 9; Conservative 5; Mismatches 6; Indels 6; Gaps 0;

QY 1 TGAQTIKGOKLYFKANGQOV 20
Db 134 TGLINVETEKLSFGANGKV 153

RESULT 12
US-11-212-443-107
; Sequence 107, Application US/11/212,443
; Publication No. US20050287165A1
; GENERAL INFORMATION:
; APPLICANT: Scalato, Enzo
; APPLICANT: Masignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Meningococcal Antigens
; FILE REFERENCE: CHIRO159
; CURRENT APPLICATION NUMBER: US/11/212,443
; CURRENT FILING DATE: 2005-08-24
; PRIOR APPLICATION NUMBER: US/09/302,626
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/IB99/00103
; PRIOR FILING DATE: 1999-01-14
; NUMBER OF SEQ ID NOS: 195
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 107
; TYPE: PRT
; LENGTH: 592
; ORGANISM: Neisseria meningitidis
US-11-212-443-107

Query Match      38.9%; Score 44; DB 7; Length 592;
Best Local Similarity 45.0%; Pred. No. 17;
Matches 9; Conservative 5; Mismatches 6; Indels 6; Gaps 0;

QY 1 TGAQTIKGOKLYFKANGQOV 20
Db 134 TGLINVETEKLSFGANGKV 153

RESULT 13
US-11-212-443-110
; Sequence 110, Application US/11/212,443
; Publication No. US20050287165A1
; GENERAL INFORMATION:
; APPLICANT: Scalato, Enzo
; APPLICANT: Masignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Meningococcal Antigens
; FILE REFERENCE: CHIRO159
; CURRENT APPLICATION NUMBER: US/11/212,443
; CURRENT FILING DATE: 2005-08-24
; PRIOR APPLICATION NUMBER: US/09/302,626
; PRIOR FILING DATE: 2005-08-24
; PRIOR APPLICATION NUMBER: PCT/IB99/00103
; PRIOR FILING DATE: 1999-01-14
; NUMBER OF SEQ ID NOS: 195
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 108
; TYPE: PRT
; LENGTH: 592
; ORGANISM: Neisseria meningitidis
US-11-212-443-110

Query Match      38.9%; Score 44; DB 7; Length 592;
Best Local Similarity 45.0%; Pred. No. 17;
Matches 9; Conservative 5; Mismatches 6; Indels 6; Gaps 0;

QY 1 TGAQTIKGOKLYFKANGQOV 20
Db 134 TGLINVETEKLSFGANGKV 153

RESULT 14
US-11-212-443-111
; Sequence 111, Application US/11/212,443
; Publication No. US20050287165A1
; GENERAL INFORMATION:
; APPLICANT: Scalato, Enzo
; APPLICANT: Masignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Meningococcal Antigens
; FILE REFERENCE: CHIRO159
; CURRENT APPLICATION NUMBER: US/11/212,443
; CURRENT FILING DATE: 2005-08-24
; PRIOR APPLICATION NUMBER: US/09/302,626
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: PCT/IB99/00103
; PRIOR FILING DATE: 1999-01-14
; NUMBER OF SEQ ID NOS: 195
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 111
; TYPE: PRT
; LENGTH: 600
; ORGANISM: Neisseria meningitidis
US-11-212-443-111

Query Match      38.9%; Score 44; DB 7; Length 600;
Best Local Similarity 45.0%; Pred. No. 17;
Matches 9; Conservative 5; Mismatches 6; Indels 6; Gaps 0;

QY 1 TGAQTIKGOKLYFKANGQOV 20
Db 142 TGLINVETEKLSFGANGKV 161

RESULT 15
US-11-067-811-1
; Sequence 1, Application US/11/067,811
; Publication No. US20050260688A1
; GENERAL INFORMATION:
; APPLICANT: Morgan, Bruce A
; APPLICANT: Enshell-Sciffers, David
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR HAIR GROWTH
; FILE REFERENCE: 10287-083001
; CURRENT APPLICATION NUMBER: US/11/067,811
; CURRENT FILING DATE: 2005-02-28
; PRIOR APPLICATION NUMBER: US 60/548,272
; PRIOR FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 1
; LENGTH: 1042
; TYPE: PRT
```

; ORGANISM: Homo sapiens  
US-11-067-811-1

US-11-067-811-1

Query Match 38.9%; Score 44; DB 7; Length 1042;

Best Local Similarity 63.6%; Pred. No. 31;

```

Local Similarity: 83.3%; 100% ID: 81.1%
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

```

Qy 9 QKLYFKANGQQ 19

[illegible]

Db 73 QKVYFKSNGSE 83

Search completed: February 10, 2006, 23:25:59  
Job time : 1.95363.secs

Job time : 1.95363 .secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:05:33 ; Search time 15.196 Seconds  
(without alignments)  
607.053 Million cell updates/sec

Title: US-10-797-821-24  
Perfect score: 104  
Sequence: 1 DANFDSIRVDVNDVADLLQ 21

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_21.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1990s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*  
9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	100.0	21	7	ADD93644
2	104	100.0	21	9	AD937267
3	104	100.0	22	2	AAR43694
4	104	100.0	22	2	AAW34158
5	104	100.0	22	7	ADD93645
6	104	100.0	22	9	AD937268
7	104	100.0	24	2	AAW34163
8	104	100.0	1017	5	AAU79285
9	104	100.0	1375	5	AAU98028
10	104	100.0	1375	5	AAU79288
11	104	100.0	1375	7	ADD93655
12	104	100.0	1375	9	AD937278
13	104	100.0	1475	5	AAU98036
14	104	100.0	1475	5	AAU98037
15	104	100.0	1475	5	AAU98040
16	104	100.0	1475	5	AAU98033
17	104	100.0	1475	5	AAU98039
18	104	100.0	1475	5	AAU98027
19	104	100.0	1475	5	AAU98032
20	104	100.0	1475	7	ADD93654
21	104	100.0	1475	9	AD937277
22	104	100.0	1476	5	AAU79284
23	104	100.0	1499	7	ADC54806
24	104	100.0	1590	7	ADD93657

25	104	100.0	1590	9	AD937280	Adx37280 Streptoco
26	103	99.0	1475	5	AAU98030	Aau98030 S. mutans
27	100	96.2	1149	6	ABR63236	Abf63236 Glucansuc
28	100	96.2	1527	5	AAU80055	Aau80055 Leuconost
29	100	96.2	1527	7	ADC54807	Adc54807 Leuconost
30	99	95.2	221	6	ABR63229	Abf63229 Glucansuc
31	99	95.2	1475	5	AAU98031	Aau98031 S. mutans
32	99	95.2	1475	5	AAU98035	Aau98035 S. mutans
33	99	95.2	1475	5	AAU98034	Aau98034 S. mutans
34	99	95.2	2022	6	ABR63232	Abf63232 Glucansuc
35	98	94.2	21	5	ABB98632	Abb98632 Dextrane-
36	98	94.2	21	5	ABB98633	Abb98633 Dextrane-
37	98	94.2	23	2	AA43346	Aay43346 GTF antiq
38	98	94.2	23	5	AAG80740	Aag80740 S. mutans
39	98	94.2	23	5	AAG80739	Aag80739 S. mutans
40	98	94.2	1475	5	AAU98038	Aau98038 S. mutans
41	98	94.2	1518	7	ADD93660	Add93660 Streptoco
42	98	94.2	1518	9	AD937283	Adx37283 Streptoco
43	95	91.3	1430	5	AAU98043	Aau98043 S. mutans
44	95	91.3	1430	5	AAU98044	Aau98044 S. mutans
45	95	91.3	1430	5	AAU98045	Aau98045 S. mutans

ALIGNMENTS

RESULT 1  
ADD93644  
ID ADD93644 standard; peptide; 21 AA.  
XX  
AC ADD93644;  
XX  
DT 29-JAN-2004 (first entry)  
XX  
DE Streptococcus mutans glucosyltransferase-B peptide fragment.  
XX  
KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.  
XX  
OS Streptococcus mutans.  
XX  
PN WO2003075845-A2.  
XX  
PD 18-SEP-2003.  
XX  
PF 07-MAR-2003; 2003WO-US006962.  
XX  
PR 07-MAR-2002; 2002US-0363209P.  
PR 08-AUG-2002; 2002US-0402483P.  
XX  
(FORS-) FORSYTH INST.  
Smith DJ, Taubman MA;  
WPI; 2003-845091/78.  
Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B (GbpB) covalently linked with a peptide subunit class II protein.

Claim 17; Page 17; 49pp; English.

The present sequence is that of a glucan-binding peptide comprising amino acid residues 442-462 of Streptococcus mutans glucosyltransferase. The peptide can be used in immunogenic compositions and subunit vaccines for dental caries. These compositions comprise a major histocompatibility complex (MHC) class II protein-binding peptide from Streptococcus mutans glucan binding protein-B (GbpB) covalently linked with a peptide subunit (such as the present peptide) of a streptococcal glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diepitopic or multiepitopic polypeptides can be prepared synthetically or by recombinant DNA technology. Antibodies raised against MHC class II binding fragments of GbpB can be used in passive immunisation.

```

XX
SQ Sequence 21 AA;
Query Match 100.0%; Score 104; DB 7; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DANFDSIRVDVNDVADLLQ 21
|||||
Db 1 DANFDSIRVDVNDVADLLQ 21
|||||

RESULT 2
ADX37267
ID ADX37267 standard; peptide; 21 AA.
XX
AC ADX37267;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B peptide #24.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
PA (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
PI Smith DJ, Taubman MA;
XX
WPI; 2005-151644/16.
XX
New composition comprising a fragment of a glucan binding protein-B
(GbpB) that binds to MHC class II protein, and a biocompatible
microparticle, useful for producing an antibody (claimed) for immunizing
mammals against dental caries.
XX
Claim 8; SEQ ID NO 24; 73pp; English.
XX
The invention relates to a composition comprising a fragment of a glucan
binding protein-B (GbpB) and a biocompatible microparticle, where the
fragment binds to a major histocompatibility complex (MHC) class II
protein. The composition is useful for producing an antibody for
immunizing mammals against dental caries. This sequence corresponds to a
Streptococcus mutans GbpB-derived peptide of the invention.
XX
Sequence 21 AA;
Query Match 100.0%; Score 104; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DANFDSIRVDVNDVADLLQ 21
|||||
Db 1 DANFDSIRVDVNDVADLLQ 21
|||||

RESULT 3
Query Match 100.0%; Score 104; DB 9; Length 21;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DANFDSIRVDVNDVADLLQ 21
|||||
Db 1 DANFDSIRVDVNDVADLLQ 21
|||||

RESULT 4
AAW34158
ID AAW34158 standard; peptide; 22 AA.
XX
AC AAW34158;
XX
DT 25-MAR-2003 (revised)
DT 18-FEB-1998 (first entry)
XX
DE GTF antigenic peptide #3.
XX
KW Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;
KW surface domain; glucan-binding domain; mutans streptococcal strain;
KW immune response; glucan-binding activity; dental caries prevention.
XX
OS Streptococcus mutans.
XX
PN US5686075-A.
XX

```

```

AAR43694
ID AAR43694 standard; peptide; 22 AA.
XX
AC AAR43694;
XX
DT 25-MAR-2003 (revised)
DT 20-MAY-1994 (first entry)
XX
DE GT subsequence.
XX
KW GT; glucosyltransferase; vaccine; T-cell; B-cell; reaction;
KW immunoresponse; peptidyl core matrix; dental caries; diphtheria; tetanus;
KW measles; polio.
XX
OS Synthetic.
XX
PN WO9322341-A1.
XX
PD 11-NOV-1993.
XX
PF 30-APR-1993; 93WO-US004094.
XX
PR 01-MAY-1992; 92US-00877295.
XX
PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.
XX
PI Taubman MA, Smith DJ;
XX
WPI; 1993-368721/46.
XX
Synthetic glucosyl-transferase peptide(s) - provoke T- and B-cell
reactions in mammals, and are effective against dental caries.
XX
Claim 3; Page 23; 38pp; English.
XX
The sequences (AAR43694-98) are subsequences from GT, they provoke T- and
B-cell responses in mammals. The can be used to produce a vaccine
comprising 2 of these sequences attached to a peptidyl core matrix. They
are useful in producing T-cell responses to the GT enzyme causing a
reduction of colonisation or accumulation of mutans streptococcal strains
in mammals. The vaccines can be used in preventing dental caries.
XX
(Updated on 25-MAR-2003 to correct PN field.)
XX
Sequence 22 AA;
Query Match 100.0%; Score 104; DB 2; Length 22;
Best Local Similarity 100.0%; Pred. No. 1.2e-09;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DANFDSIRVDVNDVADLLQ 21
|||||
Db 1 DANFDSIRVDVNDVADLLQ 21
|||||

RESULT 4
AAW34158
ID AAW34158 standard; peptide; 22 AA.
XX
AC AAW34158;
XX
DT 25-MAR-2003 (revised)
DT 18-FEB-1998 (first entry)
XX
DE GTF antigenic peptide #3.
XX
KW Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;
KW surface domain; glucan-binding domain; mutans streptococcal strain;
KW immune response; glucan-binding activity; dental caries prevention.
XX
OS Streptococcus mutans.
XX
PN US5686075-A.
XX

```



PD 11-NOV-1997.  
 XX  
 PF 30-APR-1993; 93US-000571162.  
 XX  
 PR 01-MAY-1992; 92US-00877295.  
 XX  
 PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.  
 XX  
 XX Smith DJ, Taubman MA;  
 XX WPI; 1997-558089/51.  
 XX  
 XX Immunogenic compositions containing streptococcal glucosyltransferase  
 PT peptide(s) - used for provoking immune response to streptococcal  
 PT glucosyltransferase for preventing dental caries.  
 XX  
 PS Claim 2; Col 13; 11pp; English.  
 XX  
 CC AAW34156-W34160 represent immunogenic fragments of the Streptomyces  
 CC mutans glucosyltransferase (GTF) enzyme. AAW34157 and AAW34158 are from  
 CC the catalytic domain of GTF, while AAW34160 is from the GTF surface  
 CC domain. AAW34156 and AAW34159 are from the glucan-binding domain of GTF.  
 CC These sequences can all be used in the immunogenic composition of the  
 CC invention. The composition of the invention can alternatively comprise  
 CC one or more of these sequences linked to a lysine core matrix (see  
 CC AAW34161-W34165). A composition comprising one of these sequences can be  
 CC administered to a mammal to raise an immune response, in a method for  
 CC interfering with the enzymatic activity of streptococcal  
 CC glucosyltransferase in a mammal. The immune response results in reduction  
 CC of the colonisation or accumulation of mutans streptococcal strains in  
 CC the mammal. Compositions containing AAW34156 specifically interfere with  
 CC the glucan-binding activity of the streptococcal glucosyltransferase. The  
 CC peptides can also be used in vaccines for preventing dental caries in  
 CC mammals. (Updated on 25-MAR-2003 to correct PF field.)  
 XX  
 SQ Sequence 22 AA;  
 Query Match 100.0%; Score 104; DB 2; Length 22;  
 Best Local Similarity 100.0%; Pred. NO. 1.2e-09;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DANFDSIRVDVNDVADLLQ 21  
 Db 1 DANFDSIRVDVNDVADLLQ 21  
 |||||  
 RESULT 5  
 ADD93645  
 ID ADD93645 standard; peptide; 22 AA.  
 XX  
 AC ADD93645;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Streptococcus mutans glucosyltransferase-B catalytic domain peptide.  
 XX  
 KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PN W02003075845-A2.  
 XX  
 PD 18-SEP-2003.  
 XX  
 PF 07-MAR-2003; 2003WO-US006962.  
 XX  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 XX  
 PA (FORS-) FORSYTH INST.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX WPI

DR WPI; 2003-845091/78.  
 XX  
 PT Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.  
 XX  
 XX Claim 17; Page 17; 49pp; English.  
 XX  
 CC The present sequence is that of a catalytic domain peptide fragment of  
 CC Streptococcus mutans glucosyltransferase. The peptide can be used in  
 CC immunogenic compositions and subunit vaccines for dental caries. These  
 CC compositions comprise a major histocompatibility complex (MHC) class II  
 CC protein-binding peptide from Streptococcus mutans glucan binding protein  
 CC -B (GbpB) covalently linked with a (preferably catalytic domain) peptide  
 CC subunit of a streptococcal glucosyltransferase. The compositions are used  
 CC in a claimed method of eliciting production of an antibody in a mammal.  
 CC Dieptopic or multiptopic polypeptides can be prepared synthetically or  
 CC by recombinant DNA technology. Antibodies raised against MHC class II  
 CC binding fragments of GbpB can be used in passive immunisation.  
 XX  
 SQ Sequence 22 AA;  
 Query Match 100.0%; Score 104; DB 7; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-09;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DANFDSIRVDVNDVADLLQ 21  
 Db 1 DANFDSIRVDVNDVADLLQ 21  
 |||||  
 RESULT 6  
 ADX37268  
 ID ADX37268 standard; peptide; 22 AA.  
 XX  
 AC ADX37268;  
 XX  
 DT 21-APR-2005 (first entry)  
 XX  
 DE Streptococcus mutant glucan binding protein B peptide #25.  
 XX  
 KW immunogenicity; immune stimulation; glucan binding protein-B;  
 KW microparticle; major histocompatibility complex; tooth disease.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PN US2005031633-A1.  
 XX  
 PD 10-FEB-2005.  
 XX  
 PF 09-MAR-2004; 2004US-00797821.  
 XX  
 PR 13-APR-1998; 98US-0081550P.  
 PR 08-JAN-1999; 99US-0115142P.  
 PR 12-APR-1999; 99US-00290049.  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 PR 07-MAR-2003; 2003US-00383930.  
 XX  
 XX (SMIT/) SMITH D J.  
 PA (TAUB/) TAUBMAN M A.  
 XX  
 PI Smith DJ, Taubman MA;  
 XX WPI; 2005-151644/16.  
 XX  
 XX New composition comprising a fragment of a glucan binding protein-B  
 PT (GbpB) that binds to MHC class II protein, and a biocompatible  
 PT microparticle, useful for producing an antibody (claimed) for immunizing  
 PT mammals against dental caries.  
 XX  
 XX Claim 8; SEQ ID NO 25; 73pp; English.  
 XX

CC The invention relates to a composition comprising a fragment of a glucan  
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II  
 CC protein. The composition is useful for producing an antibody for  
 CC immunizing mammals against dental caries. This sequence corresponds to a  
 CC Streptococcus mutans GbpB-derived peptide of the invention.  
 XX  
 SQ Sequence 22 AA;

Query Match 100.0%; Score 104; DB 9; Length 22;  
 Best Local Similarity 100.0%; Pred. No. 1.2e-09;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQ 21  
 |||||

Db 1 DANFDSIRVDVNDVNDADLLQ 21

# RESULT 7

AAW34163  
 ID AAW34163 standard; peptide; 24 AA.

XX AC AAW34163;

XX DT 25-MAR-2003 (revised)

XX DT 18-FEB-1998 (first entry)

XX DE GTF antigenic peptide #3 linked to polylysine core.

XX KW Glucosyltransferase; GTF; enzyme; immunogen; catalytic domain; vaccine;  
 KW surface domain; glucan-binding domain; mutans streptococcal strain;  
 KW immune response; glucan-binding activity; dental caries prevention.  
 XX

OS Synthetic.

OS OS Streptococcus mutans.

XX FH Key Location/Qualifiers

XX FT Peptide 1..22

XX FT /label= GTF antigenic peptide #3 (see AAW34158)  
 FT /note= "attached to the dendritic polylysine core via the  
 FT alpha-amino group of Lys(23); a second copy of the  
 FT antigenic 22-mer is linked to Lys(23) via the omega amino  
 FT group"  
 XX

XX FT Modified-site 23

XX FT /note= "Lys(23) is linked to one copy of the antigenic  
 FT peptide through the alpha-amino group, and to a second  
 FT copy of the peptide (not shown) via the omega amino  
 FT group"  
 XX

XX FT Modified-site 24

XX FT /note= "the alpha amino acid group of Lys(24) forms a  
 FT peptide linkage with the carboxyl amino group of Lys(23);  
 FT the omega amino group of Lys(24) forms a peptide bond  
 FT with a second Lys residue analogous to Lys(23)"  
 XX

XX PN US5686075-A.

XX PD 11-NOV-1997.

XX PF 30-APR-1993; 93US-00057162.

XX PR 01-MAY-1992; 92US-00877295.

XX PA (FORS-) FORSYTH DENTAL INFIRMARY FOR CHILDREN.

XX PI Smith DJ, Taubman MA;

XX PR WPI; 1997-558089/51.

XX PT Immunogenic compositions containing streptococcal glucosyl:transferase  
 PT peptide(s) - used for provoking immune response to streptococcal  
 PT glucosyl:transferase for preventing dental caries.  
 XX

XX PS Claim 12; Col; 11pp; English.

XX CC AAW34161-W34165 represent the Streptomyces mutans glucosyltransferase  
 CC (GTF) enzyme immunogenic fragments shown in AAW34156-W34160 linked to a  
 CC polylysine core. AAW34157 and AAW34158 are from the catalytic domain of  
 CC GTF, while AAW34160 is from the GTF surface domain. AAW34156 and AAW34159  
 CC are from the glucan-binding domain of GTF. These sequences, and the  
 CC immunogenic fragments shown in AAW34156-W34160 can all be used in the  
 CC immunogenic composition of the invention. A composition comprising one of  
 CC these sequences can be administered to a mammal to raise an immune  
 CC response, in a method for interfering with the enzymatic activity of  
 CC streptococcal glucosyltransferase in a mammal. The immune response  
 CC results in reduction of the colonisation or accumulation of mutans  
 CC streptococcal strains in the mammal. Compositions containing AAW34156  
 CC specifically interfere with the glucan-binding activity of the  
 CC streptococcal glucosyltransferase. The peptides can also be used in  
 CC vaccines for preventing dental caries in mammals. (Updated on 25-MAR-2003  
 CC to correct PF field.)  
 XX  
 SQ Sequence 24 AA;

Query Match 100.0%; Score 104; DB 2; Length 24;

Best Local Similarity 100.0%; Pred. No. 1.3e-09;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQ 21  
 |||||

Db 1 DANFDSIRVDVNDVNDADLLQ 21

# RESULT 8

AAU79285

ID AAU79285 standard; protein; 1017 AA.

XX AC AAU79285;

XX DT 13-AUG-2002 (first entry)

XX DE Streptococcus mutans monoclonal antibody-related protein #2.

XX KW Antibody; dental caries; water insoluble glucan synthetase; anti-carries;  
 KW Glucosyl transferase-B; immunotherapy.  
 XX

OS Streptococcus mutans.

XX PN JP2002114709-A.

XX PD 16-APR-2002.

XX PF 04-OCT-2000; 2000JP-00304889.

XX PR 04-OCT-2000; 2000JP-00304889.

XX PA (UYNI-) UNIV NIPPON.

XX XX WPI; 2002-448885/48.

XX PT Anti-carries agent composed of a monoclonal antibody against an inhibitory  
 PT enzyme against water insoluble glucan synthetase of glucosyl transferase-  
 PT B (GTF-B) of Streptococcus mutans.  
 XX

XX PS Claim 4; Page 17-19; 28pp; Japanese.

XX CC The invention relates to a monoclonal antibody against dental caries and  
 CC an anti-carries agent composed of a monoclonal antibody produced by  
 CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)  
 CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having  
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
 CC transferase-B. The monoclonal antibody specifically inhibits water  
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl  
 CC transferase-B and is used in the immunotherapy of dental caries. This  
 CC sequence represents a Streptococcus mutans monoclonal antibody-related  
 CC protein  
 XX

SQ Sequence 1017 AA;

Query Match 100.0%; Score 104; DB 5; Length 1017;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVDADLLQ 21  
 Db 408 DANFDSIRVDVNDVDADLLQ 428

RESULT 9  
 AAU98028  
 ID AAU98028 standard; protein; 1375 AA.  
 XX AC AAU98028;  
 XX  
 DT 27-AUG-2002 (first entry)  
 XX  
 DE S. mutans Glucosyltransferase GTFC.  
 XX  
 KW Glucosyltransferase; GTFC; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PN US2002031826-A1.  
 XX  
 PD 14-MAR-2002.  
 XX  
 PF 19-DEC-2000; 2000US-00740274.  
 XX  
 PR 07-JUN-1995; 95US-00478704.  
 PR 07-JUN-1995; 95US-00482711.  
 PR 07-JUN-1995; 95US-00485243.  
 PR 16-JAN-1998; 98US-00007999.  
 PR 16-JAN-1998; 98US-00008172.  
 PR 20-JAN-1998; 98US-00009620.  
 PR 11-DEC-1998; 98US-00210361.  
 XX  
 PA (NICH/) NICHOLS S E.  
 XX  
 PI Nichols SE;  
 XX  
 DR WPI; 2002-414332/44.  
 DR N-PSDB; ABK52939.  
 XX  
 XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.  
 XX  
 PS Disclosure; Page 30-33; 44pp; English.  
 XX

The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful

CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents GTFC

SQ Sequence 1375 AA;

Query Match 100.0%; Score 104; DB 5; Length 1375;  
 Best Local Similarity 100.0%; Pred. No. 1.9e-07;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVDADLLQ 21  
 Db 468 DANFDSIRVDVNDVDADLLQ 488

RESULT 10  
 AAU79288  
 ID AAU79288 standard; protein; 1375 AA.  
 XX AAU79288;  
 XX  
 DT 13-AUG-2002 (first entry)  
 DE Streptococcus mutans monoclonal antibody-related protein #5.  
 XX  
 KW Antibody; dental caries; water insoluble glucan synthetase; anti-carries;  
 KW glucosyl transferase-B; immunotherapy.  
 XX  
 OS Streptococcus mutans.  
 XX  
 PN JP2002114709-A.  
 XX  
 PD 16-APR-2002.  
 XX  
 PF 04-OCT-2000; 2000JP-00304889.  
 XX  
 PR 04-OCT-2000; 2000JP-00304889.  
 XX  
 PA (UYNI-) UNIV NIPPON.  
 XX  
 XX WPI; 2002-448885/48.  
 DR  
 XX  
 PT Anti-carries agent composed of a monoclonal antibody against an inhibitory  
 PT enzyme against water insoluble glucan synthetase of glucosyl transferase-  
 PT B (GTF-B) of Streptococcus mutans.  
 XX  
 PS Disclosure; Page 22-25; 28pp; Japanese.  
 XX

The invention relates to a monoclonal antibody against dental caries and  
 CC an anti-carries agent composed of a monoclonal antibody produced by  
 CC Streptococcus mutans, particularly mouse-hybridoma MHP126 (FERM P-17566)  
 CC or mouse-hybridoma MHP136 (FERM P-17567), against an enzyme having  
 CC inhibitive activity against water insoluble glucan synthetase of glucosyl  
 CC transferase-B. The monoclonal antibody specifically inhibits water  
 CC insoluble glucan synthetase of Streptococcus mutans produced glucosyl  
 CC transferase-B and is used in the immunotherapy of dental caries. This  
 CC sequence represents a Streptococcus mutans monoclonal antibody-related  
 CC protein  
 XX  
 SQ Sequence 1375 AA;

```
Query Match      100.0%; Score 104; DB 5; Length 1375;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DANFDSIRVDVNDVADLLQ 21
Db      468 DANFDSIRVDVNDVADLLQ 488

RESULT 11
ADD93655
ID ADD93655 standard; protein; 1375 AA.
XX
AC ADD93655;
XX
DT 29-JAN-2004 (first entry)
XX
DE Streptococcus mutans glucosyltransferase-C.
XX
KW Glucosyltransferase; enzyme; vaccine; anticaries; epitope; immunogen.
XX
OS Streptococcus mutans.
XX
PN WO2003075845-A2.
XX
PD 18-SEP-2003.
XX
PF 07-MAR-2003; 2003WO-US006962.
XX
PR 07-MAR-2002; 2002US-0363209P.
XX
PR 08-AUG-2002; 2002US-0402483P.
XX
PA (FORS-) FORSYTH INST.
XX
PI Smith DJ, Taubman MA;
XX
WPI; 2003-845091/78.
XX
Composition useful as vaccines for dental caries comprises a fragment of
PT a glucan binding protein-B binding to a major histocompatibility complex
PT class II protein.
XX
Claim 16; Page 13; 49pp; English.
XX
The present sequence is the protein sequence of Streptococcus mutans
CC Glucosyltransferase-C (GTF-C). Peptide fragments of GTF-C, especially
CC from the catalytic domain of the polypeptide, can be used in immunogenic
CC compositions and subunit vaccines for dental caries. These compositions
CC comprise a major histocompatibility complex (MHC) class II protein-
CC binding peptide from S. mutans glucan binding protein-B (GbpB)
CC covalently linked with a peptide fragment of a streptococcal
CC glucosyltransferase. The compositions are used in a claimed method of
CC eliciting production of an antibody in a mammal. Diepitopic or
CC multiepitopic polypeptides can be prepared synthetically or by
CC recombinant DNA technology. Antibodies raised against MHC class II
CC binding fragments of GbpB can be used in passive immunisation.
XX
SQ Sequence 1375 AA;

Query Match      100.0%; Score 104; DB 7; Length 1375;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DANFDSIRVDVNDVADLLQ 21
Db      468 DANFDSIRVDVNDVADLLQ 488

RESULT 12
ADX37278
ID ADX37278 standard; protein; 1375 AA.
XX
```

```
AC ADX37278;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutans glucan binding protein B variant #7.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
(SMIT/) SMITH D J.
(PAUB/) TAUBMAN M A.
XX
PI Smith DJ, Taubman MA;
XX
WPI; 2005-151644/16.
XX
New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
Claim 7; SEQ ID NO 35; 73pp; English.
XX
The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
SQ Sequence 1375 AA;

Query Match      100.0%; Score 104; DB 9; Length 1375;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DANFDSIRVDVNDVADLLQ 21
Db      468 DANFDSIRVDVNDVADLLQ 488

RESULT 13
AAU98036
ID AAU98036 standard; protein; 1475 AA.
XX
AC AAU98036;
XX
DT 27-AUG-2002 (first entry)
XX
DE S. mutans glucosyltransferase GTFB mutant D567T/D571K.
XX
KW Glucosyltransferase; GTFB; transgenic plant; paper sizing;
KW coating composition; glucan; starch; latex; thermoplastic molecule;
KW amyloplast; vacuole; paper manufacture; mutant; mutein.
XX
OS Streptococcus mutans.
XX
XX Synthetic.
```



CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36

XX Sequence 1475 AA;

Query Match 100.0%; Score 104; DB 5; Length 1475;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-07;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQ 21  
 Db |||||||||||||||||||||

442 DANFDSIRVDVNDVADLLQ 462

RESULT 15

AAU98040

ID AAU98040 standard; protein; 1475 AA.

AC AAU98040;

DT 27-AUG-2002 (first entry)

DE S. mutans glucosyltransferase GTFB mutant K779Q.

XX Glucosyltransferase; GTFB; transgenic plant; paper sizing;  
 KW coating composition; glucan; starch; latex; thermoplastic molecule;  
 KW amyloplast; vacuole; paper manufacture; mutant; mutin.

OS Streptococcus mutans.

OS Synthetic.

OS Key Location/Qualifiers

FT Key Misc-difference 779 /note= "Wild-type Lys substituted by Gln"

FT US2002031826-A1.

PN 14-MAR-2002.

PD 19-DEC-2000; 2000US-00740274.

PF 07-JUN-1995; 95US-00478704.

PR 07-JUN-1995; 95US-00482711.

PR 07-JUN-1995; 95US-00485243.

PR 16-JAN-1998; 98US-00007999.

PR 16-JAN-1998; 98US-00008172.

PR 20-JAN-1998; 98US-00009620.

PR 11-DEC-1998; 98US-00210361.

XX (NICH/) NICHOLS S E.

XX Nichols SE;  
 PI WPI; 2002-414332/44.

XX Glucosyltransferase B or D protein useful for producing a glucan useful  
 PT as substitutes for and additions to modified starch and latexes in paper  
 PT manufacture, comprises mutations in specific positions.

XX Claim 36; Page; 44pp; English.

XX The invention an isolated protein comprising a glucosyltransferase (GTF)  
 CC B polypeptide having changes at position from I448V, D457N, D567T,  
 CC K1014T, D457N/D567T, D457N/D571K, D567T/D571K, D567T/D571K/K1014T,  
 CC I448V/D457N/D567T/D571K/K779Q/K1014T, Y169A/Y170A/Y171A, and K779Q, or a  
 CC GTF D polypeptide having changes at positions from T589D, T589E, N471D,  
 CC N471D/T589D, and N471D/T589E. Also included are a glucan produced by the  
 CC GTF mutant, an isolated polynucleotide which encodes P1 or P2, or its  
 CC complementary polynucleotide, a ribonucleic acid sequence encoding the  
 CC GTF mutant, an expression cassette comprising the polynucleotide operably  
 CC linked to a promoter, a vector comprising the expression cassette, host  
 CC cell introduced with the vector, a transgenic plant comprising the  
 CC vector, a seed or tuber from the transgenic plant, a paper sizing and/or  
 CC coating composition comprising a glucan produced in a plant transformed  
 CC with a gene encoding the mutant GTF, wild type or, starch, a latex,  
 CC thermoplastic molecule or their combinations or glucan and starch where  
 CC the glucan is produced in the amyloplast and/or vacuole or a maize line  
 CC deficient in starch biosynthesis, transformed with a gene encoding a  
 CC glucosyltransferase B or D enzyme, wild-type or mutant and a paper  
 CC comprising the glucan (paper sizing/coating agent). The vector is useful  
 CC for producing a glucan in a plant. The method comprises transforming a  
 CC plant cell with the vector, growing the plant cell under plant growing  
 CC conditions to produce a regenerated plant and inducing expression of the  
 CC polynucleotide for a time sufficient to produce the glucan in the  
 CC regenerated plant, where the vector contains a transit sequence from  
 CC ribulose biphosphate carboxylase small subunit, waxy, brittle-1 and  
 CC chlorophyll AB binding protein to produce a transgenic plant, and glucan  
 CC is produced in the amyloplast of potato or the vacuole of sugar beet.  
 CC Glucans are useful as substitutes for and additions to modified starch  
 CC and latexes in paper manufacture. Unlike prior art techniques, which  
 CC require input materials that produce chemical effluents, paper  
 CC manufacture utilising the glucan produced by GTF, which utilises  
 CC biologically produced input materials, is more cost-effective and  
 CC environmentally friendly. Moreover, glucans also exhibit thermoplastic  
 CC properties and impart gloss to the paper during coating step. The present  
 CC sequence represents a GTFB mutant of the invention. Note: The present  
 CC sequence is not shown in the specification but was created by the indexer  
 CC using the GTFB sequence appearing as AAU98027 and the information in  
 CC claim 36

XX Sequence 1475 AA;

Query Match 100.0%; Score 104; DB 5; Length 1475;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-07;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQ 21  
 Db |||||||||||||||||||||

442 DANFDSIRVDVNDVADLLQ 462

Search completed: February 10, 2006, 22:19:42

Job time : 16.1996 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:20:08 ; Search time 2.47681 Seconds  
(without alignments)  
815.787 Million cell updates/sec

Title: US-10-797-821-24  
Perfect score: 104  
Sequence: 1 DANFDSIRVDVNDVADLLQ 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 80.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	100.0	1375	2 JT0345	dextranucrase (EC
2	104	100.0	1475	2 B33135	gtfB protein precu
3	98	94.2	1518	2 A44811	glucosyltransferas
4	94	90.4	1365	2 A41483	glucosyltransferas
5	94	90.4	1508	2 T31098	probable dextranu
6	92	88.5	1577	2 T30858	glucosyltransferas
7	92	88.5	1592	2 A38175	glucosyltransferas
8	90	86.5	1431	2 A45866	dextranucrase (EC
9	89	85.6	1290	2 JC5473	dextranucrase (EC
10	89	85.6	1449	2 T30857	glucosyltransferas
11	89	85.6	1449	2 T30552	glucosyltransferas
12	83	79.8	1599	2 S22737	glucosyltransferas
13	54	51.9	655	1 ALKBG	cyclomaltodextrin
14	46	44.2	673	2 B70528	probable peptidase
15	45.5	43.8	147	2 D69732	PSX prophage ORF
16	45.5	43.8	378	1 A40004	histidine decarbox
17	45.5	43.8	512	2 S61905	threonine synthase
18	45	43.3	457	2 G82925	hypothetical prote
19	44	42.3	128	2 AC1607	transcription term
20	44	42.3	280	2 E72296	pantoate-beta-alan
21	44	42.3	315	2 AF1393	glycosyl transfera
22	44	42.3	315	2 A11768	glycosyl transfera
23	44	42.3	1819	2 D97033	uncharacterized pr
24	43.5	41.8	550	2 A82378	conserved hypothet
25	43.5	41.8	762	2 C69657	cobalamin-independ
26	43	41.3	9	2 A39841	sucrose 3-glucosyl
27	43	41.3	307	2 D84536	hypothetical prote
28	43	41.3	352	2 G95895	probable mechyltra
29	43	41.3	385	1 S48456	histidinol-phospha

30	43	41.3	448	2 B72262	glucose-6-phosphat
31	43	41.3	451	2 F97230	probable maltodext
32	43	41.3	594	2 S37816	hypothetical prote
33	43	41.3	638	2 S41159	sodium transport p
34	43	41.3	668	2 S49639	probable membrane
35	43	41.3	826	2 T06494	1,4-alpha-glucan b
36	42.5	40.9	147	2 H69947	phage-related prot
37	42.5	40.9	486	2 AG0355	probable membrane
38	42.5	40.9	1214	2 H75034	reverse gyrase (to
39	42	40.4	179	2 S35942	probable ATP synth
40	42	40.4	179	2 S48643	ATP synthase - soy
41	42	40.4	316	2 A83588	streptogramin lyas
42	42	40.4	374	2 G90100	histone deacetylase
43	42	40.4	527	2 F84256	IMP dehydrogenase
44	42	40.4	640	2 I51915	epithelial sodium
45	42	40.4	765	1 B65202	formate C-acetylitr

ALIGNMENTS

RESULT 1

JT0345  
dextranucrase (EC 2.4.1.5) precursor - Streptococcus mutans (strain GS-5)  
N;Alternate names: sucrose 6-glucosyltransferase  
C;Species: Streptococcus mutans  
C;Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 09-Jul-2004  
C;Accession: JT0345; C33135  
R;Ueda, S.; Shiroza, T.; Kuramitsu, H.K.  
Gene 69, 101-109, 1988  
A;Title: Sequence analysis of the gtfC gene from Streptococcus mutans GS-5.  
A;Reference number: JT0345; MUID:89137980; PMID:2976010  
A;Accession: JT0345  
A;Molecule type: DNA  
A;Residues: 1-1375 <UED>  
A;Cross-references: UNIPROT:P13470; UNIPARC:UPI0000155515  
A;Experimental source: GS-5  
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
J. Bacteriol. 169, 4263-4270, 1987  
A;Title: Sequence analysis of the gtfB gene from Streptococcus mutans.  
A;Reference number: A33135; MUID:87308013; PMID:3040685  
A;Accession: C33135  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-349 <SHI>  
A;Cross-references: UNIPARC:UPI000014E25D; GB:M17361  
C;Genetics:  
A;Gene: gtfC  
C;Function:  
A;Description: catalyzes the synthesis of both water-soluble and water-insoluble glucans  
C;Keywords: duplication; glycosyltransferase; hexosyltransferase  
F;1-34/Domain: signal sequence #status predicted <SIG>  
F;35-1375/Product: glucosyltransferase #status predicted <MAT>  
F;1126-1145/Domain: cpl repeat homology <CP1>  
F;1253-1272/Domain: cpl repeat homology <CP2>  
F;1318-1337/Domain: cpl repeat homology <CP3>

Query Match 100.0%; Score 104; DB 2; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 1.8e-07;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVADLLQ 21  
|||||  
Db 468 DANFDSIRVDVNDVADLLQ 488

RESULT 2

B33135  
gtfB protein precursor - Streptococcus mutans  
C;Species: Streptococcus mutans  
C;Date: 23-Oct-1990 #sequence\_revision 23-Oct-1990 #text\_change 09-Jul-2004  
C;Accession: B33135; A33128  
R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.

J. Bacteriol. 169, 4263-4270, 1987  
A;Title: Sequence analysis of the gtfA gene from Streptococcus mutans.  
A;Reference number: A33135; MUID:97308013; PMID:3040685  
A;Accession: B33135  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-1475 <SHI>  
A;Cross-references: UNIPROT:P08987; UNIPARC:UPI000014D972; GB:M17361; NID:g153639; PIDN:R;Shiroza, T.; Ueda, S.; Kuramitsu, H.K.  
Submitted to the Protein Sequence Database, September 1990  
A;Reference number: A33128  
A;Accession: A33128  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-171,173-641,'N',643-1475 <SH2>  
A;Cross-references: UNIPARC:UPI000017ACSE  
A;Experimental source: strain GS-5  
F;1096-1115/Domain: cpl repeat homology <CP1>  
F;1224-1243/Domain: cpl repeat homology <CP2>  
F;1289-1308/Domain: cpl repeat homology <CP3>  
F;1354-1373/Domain: cpl repeat homology <CP4>  
F;1419-1438/Domain: cpl repeat homology <CP5>  
Query Match 100.0%; Score 104; DB 2; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 1.9e-07;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 DANFDSIRVDVNDVNDADLLQ 21  
Db 442 DANFDSIRVDVNDVNDADLLQ 462  
RESULT 3  
A44811  
glucosyltransferase (EC 2.4.1.-) I - Streptococcus salivarius  
C;Species: Streptococcus salivarius  
C;Date: 31-Dec-1993 #sequence revision 31-Dec-1993 #text\_change 09-Jul-2004  
C;Accession: A44811; S22726; S28809  
R;Giffard, P.M.; Simpson, C.L.; Milward, C.P.; Jacques, N.A.  
J. Gen. Microbiol. 137, 2577-2593, 1991  
A;Title: Molecular characterization of a cluster of at least two glucosyltransferase genes  
A;Reference number: A44811; MUID:92148377; PMID:1838391  
A;Accession: A44811  
A;Molecule type: DNA  
A;Residues: 1-1518 <GIF>  
A;Cross-references: UNIPROT:Q00600; UNIPARC:UPI000000BEF31; EMBL:Z11873; NID:g47526; PIDN:A;Note: sequence extracted from NCBI backbone (NCBIN:81050, NCBIP:81052)  
C;Genetics:  
A;Gene: gtfJ  
C;Keywords: glycosyltransferase; hexosyltransferase  
F;1307-1326/Domain: cpl repeat homology <CP4>  
Query Match 94.2%; Score 98; DB 2; Length 1518;  
Best Local Similarity 90.5%; Pred. No. 1.6e-06;  
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
QY 1 DANFDSIRVDVNDVNDADLLQ 21  
Db 462 DANFDSIRVDVNDVNDADLLQ 482  
RESULT 4  
A41483  
glucosyltransferase (EC 2.4.1.-) gtfS precursor - Streptococcus sobrinus  
C;Species: Streptococcus sobrinus  
C;Date: 30-Jun-1992 #sequence revision 30-Jun-1992 #text\_change 15-Mar-2004  
C;Accession: A41483  
R;Gilmore, K.S.; Russell, R.R.B.; Ferretti, J.J.  
Infect. Immun. 58, 2452-2458, 1990  
A;Title: Analysis of the Streptococcus downei gtfS gene, which specifies a glucosyltransferase  
A;Reference number: A41483; MUID:90316665; PMID:2142479  
A;Accession: A41483  
A;Molecule type: DNA

A;Residues: 1-1365 <GIL>  
A;Cross-references: UNIPARC:UPI000012BCB6; GB:M30943; NID:g153652; PIDN:AAA26898.1; PID:G  
C;Genetics:  
A;Gene: gtfS  
C;Keywords: glycosyltransferase; hexosyltransferase  
Query Match 90.4%; Score 94; DB 2; Length 1365;  
Best Local Similarity 85.7%; Pred. No. 5.7e-06;  
Matches 18; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 DANFDSIRVDVNDVNDADLLQ 21  
Db 428 DANFDGVRVDVNDVNDADLLQ 448  
RESULT 5  
T31098  
probable dextranucrase (EC 2.4.1.5), extracellular - Leuconostoc mesenteroides  
C;Species: Leuconostoc mesenteroides  
C;Date: 22-Oct-1999 #sequence revision 22-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T31098  
R;Monchois, V.; Renaud-Simeon, M.; Monsan, P.; Willemot, R.M.  
FEMS Microbiol. Lett. 159, 307-315, 1998  
A;Title: Cloning and sequencing of a gene coding for an extracellular dextranucrase (DSF)  
A;Reference number: Z20981; MUID:98164374; PMID:9503626  
A;Accession: T31098  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1508 <MON>  
A;Cross-references: UNIPROT:O52224; UNIPARC:UPI000000BB69B; EMBL:AF030129; NID:g2766611; I  
A;Experimental source: strain NRRL B-1299  
C;Genetics:  
A;Gene: dsrB  
C;Function:  
A;Description: produces dextran composed only of alpha(1-6) glucosidic bonds  
C;Keywords: glycosyltransferase; hexosyltransferase  
Query Match 90.4%; Score 94; DB 2; Length 1508;  
Best Local Similarity 95.0%; Pred. No. 6.4e-06;  
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 2 ANPDSIRVDVNDVNDADLLQ 21  
Db 525 ANPDSIRVDVNDVNDADLLQ 544  
RESULT 6  
T30858  
glucosyltransferase - Streptococcus salivarius  
C;Species: Streptococcus salivarius  
C;Date: 22-Oct-1999 #sequence revision 22-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T30858  
R;Simpson, C.L.; Giffard, P.M.; Jacques, N.A.  
Infect. Immun. 63, 609-621, 1995  
A;Title: Streptococcus salivarius ATCC 25975 possesses at least two genes coding for prin  
A;Reference number: Z20909; MUID:95122197; PMID:7822030  
A;Accession: T30858  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-1577 <SIM>  
A;Cross-references: UNIPROT:Q55265; UNIPARC:UPI000000B8087; EMBL:L35928; NID:g662380; PID  
C;Genetics:  
A;Gene: gtfm  
Query Match 88.5%; Score 92; DB 2; Length 1577;  
Best Local Similarity 85.0%; Pred. No. 1.3e-05;  
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 2 ANPDSIRVDVNDVNDADLLQ 21  
Db 550 ANPDSIRVDVNDVNDADLLQ 569



dextranucrase (EC 2.4.1.5) - Leuconostoc mesenteroides  
C:Species: Leuconostoc mesenteroides  
C:Date: 07-Jul-1997 #sequence\_revision 29-Aug-1997 #text\_change 09-Jul-2004  
C:Accession: JC5473

Query Match 85.6%; Score 89; DB 2; Length 1449;  
Best Local Similarity 81.0%; Pred. NO. 3.5e-05;  
Matches 17: Conservative 2; Mismatches 2; Indels



A;Residues: 1-147 <KUN>  
A;Cross-references: UNIPROT:P54332; UNIPARC:UPI0000060282; GB:Z99110; GB:AL009126; NID:9  
A;Experimental source: strain 168  
C;Genetics:  
A;Gene: xkdm  
Query Match 43.8%; Score 45.5; DB 2; Length 147;  
Best Local Similarity 42.9%; Pred. No. 10;  
Matches 9; Conservative 7; Mismatches 4; Indels 1; Gaps 1;  
QY 1 DANFDSIRVDVNDVDADLLQ 21  
| | | | | : : : | : : : | :  
Db 108 DYNFDSAKIASLD-VDSEALE 127  
Search completed: February 10, 2006, 22:33:14  
Job time : 5.60181 secs

100 88 73 24

**THIS PAGE BLANK (USPTO)**

GN  
OS  
OC  
OC  
OC  
NCBI\_TaxID=1245;  
[1]  
RN  
RP  
RC  
RA  
RN  
RT  
RT  
RT

Name:adrrt;  
Leuconostoc mesenteroides.  
Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
TaxID=1245;  
[1]  
Nucleotide sequence.  
Strain=IBT-PQ;  
Fernandez-Vazquez J.L., Lopez-Mungua A., Olvera C.  
"Molecular characterization of a dextranucrase gene  
from mesenteroides IBT-PQ isolated from pulque.";

```

RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY504865; AN579426.1; -; Genomic DNA.
DR GO: 0047849; F:dextranase activity; IEA.
DR GO: 0016757; F:transferase activity, transferring glycosyl. .; IEA.
DR GO: 0009250; P:glucan biosynthesis; IEA.
DR InterPro: IPR002479; CW binding.
DR InterPro: IPR003318; Glyco hydro 70.
DR Pfam: PF01473; CW binding; 1.
DR Pfam: PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1454 AA; 160569 MW; 2DC7342963B6FD87 CRC64;

Query Match 100.0%; Score 104; DB 2; Length 1454;
Best Local Similarity 100.0%; Pred No. 1.4e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQ 21
Db 464 DANFDSIRVDVNDVNDADLLQ 484

RESULT 3
GTFC_STRMU
ID GTFC_STRMU STANDARD; PRT; 1455 AA.
AC P13470; O69382; O69385; O69388; O69391; O69397; P05427;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glycosyltransferase-SI precursor (EC 2.4.1.5) (GTFC-SI)
DE (Dextranucrase) (Sucrose 6-glucosyltransferase).
GN Names:gtfc; OrderedLocusNames:SMU.1005;
OS Streptococcus mutans;
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=89117980; PubMed=2976010; DOI=10.1016/0378-1119(88)90382-4;
RA Ueda S., Shiroza T., Kuramitsu H.K.;
RT "Sequence analysis of the gtfc gene from Streptococcus mutans GS-5." ;
RL Gene 69:101-109(1988).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT4239 / Serotype c, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT8148 / Serotype c;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Sobue S.,
RA Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RT Streptococcus mutans." ;
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.N., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen." ;
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
RN [4]
RP NUCLEOTIDE SEQUENCE OF 1-349.
RC STRAIN=GS-5;
RX MEDLINE=87308013; PubMed=3040685;
RA Shiroza T., Ueda S., Kuramitsu H.K.;
RT "Sequence analysis of the gtfc gene from Streptococcus mutans." ;
CC J. Bacteriol. 169:4263-4270(1987).
CC -!- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of
CC their ability to adhere to smooth surfaces and mediate the

```

```

CC aggregation of bacterial cells and food debris.
CC -!- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-
CC fructose + (1,6-alpha-D-glucosyl)(n+1).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
CC water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
CC forms of glucans.
CC -!- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -!- SIMILARITY: Contains 5 cell wall binding repeats.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC
CC EMBL: M22054; AAA88592.1; -; Genomic DNA.
CC EMBL: D88652; BAA26102.1; -; Genomic DNA.
CC EMBL: D88655; BAA26106.1; -; Genomic DNA.
CC EMBL: D88658; BAA26110.1; -; Genomic DNA.
CC EMBL: D88661; BAA26114.1; -; Genomic DNA.
CC EMBL: D89978; BAA26120.1; -; Genomic DNA.
CC EMBL: AE014940; AAN58706.1; -; Genomic DNA.
CC PIR: M17361; AAA88589.1; -; Genomic DNA.
CC HSSP: P06653; 1H8G.
CC InterPro: IPR002479; Cell_wall_bd_put.
CC InterPro: IPR003318; Glyco_hydro_70.
CC Pfam: PF01473; CW binding; 1; 2.
CC Pfam: PF02324; Glyco_hydro_70; 1.
CC Complete proteome; Dental Caries; Glycosyltransferase; Repeat; Signal;
KW Transferase.
FT SIGNAL 1 34
FT CHAIN 35 1455 Glucosyltransferase-SI.
FT REPEAT 1126 1159 A repeat.
FT REPEAT 1169 1200 A repeat.
FT REPEAT 1227 1238 C repeat.
FT REPEAT 1253 1303 AC repeat.
FT REPEAT 1318 1330 AC repeat (incomplete).
FT REGION 35 1050 Catalytic (approximate).
FT REGION 1126 1455 2.4 A, 1 C and 1 AC repeats.
FT REGION 1126 1455 Glucan-binding (approximate).
FT VARIANT 21 21 V -> I (in strain GS-5).
FT VARIANT 81 81 P -> L (in strain MT4239).
FT VARIANT 106 106 D -> V (in strain GS-5).
FT VARIANT 116 116 S -> A (in strain GS-5 and strain
FT VARIANT 126 126 A -> T (in strain GS-5).
FT VARIANT 150 151 SR -> PK (in strain GS-5, strain MT4239
FT VARIANT 256 256 A -> V (in strain GS-5 and strain
FT VARIANT 425 425 R -> N (in strain MT4251).
FT VARIANT 519 519 Y -> D (in strain MT4245 and strain
FT VARIANT 538 538 R -> K (in strain MT4245 and strain
FT VARIANT 545 545 Y -> F (in strain MT4245 and strain
FT VARIANT 597 597 N -> D (in strain MT4245, strain MT4251,
FT VARIANT 600 600 strain MT4467 and strain MT8148).
FT VARIANT 601 601 R -> K (in strain MT4245, strain MT4251,
FT VARIANT 614 614 strain MT4467 and strain MT8148).
FT VARIANT 727 727 A -> T (in strain GS-5).
FT VARIANT 734 734 M -> T (in strain GS-5).
FT VARIANT 734 734 T -> I (in strain MT8148).
FT VARIANT 964 964 A -> V (in strain MT8148).
FT VARIANT 1113 1113 L -> F (in strain MT4239).
FT VARIANT 1118 1118 N -> Y (in strain MT4239).
FT VARIANT 1204 1204 A -> T (in strain MT4239).
FT VARIANT 1204 1204 I -> V (in strain GS-5, strain MT4239,
FT strain MT4467 and strain MT8148).

```

```

FT VARIANT 1208 1209 V -> I (in strain MT8148).
FT VARIANT 1292 1294 DGH -> NGY (in strain GS-5, strain MT4467
FT VARIANT 1305 1369 and strain MT8148).
FT VARIANT 1326 1331 I -> V (in strain MT8148).
FT VARIANT 1331 1331 T -> A (in strain GS-5, strain MT4239,
FT VARIANT 1377 1377 strain MT4467 and strain MT8148).
FT VARIANT 1398 1398 R -> K (in strain MT8148).
FT VARIANT 1424 1424 V -> I (in strain MT8148).
FT VARIANT 1439 1439 D -> N (in strain MT4239).
FT VARIANT 1444 1444 V -> I (in strain MT4239 and strain
FT VARIANT 1455 1455 MT8148).
FT VARIANT 1455 1455 S -> P (in strain MT8148).
FT CONFLICT 1337 1455 QRLFKNGVQAKGELITERKRIYDNPNGNEVNRVYR
FT TSSNGWYFGNDGVALIGMHVVEGRVYFDENGVRVYASHD
FT ORNHWDYDRDFGRGSSAVFRHRSRNGFFDNFRF ->
FT HASILSLMVFRLRESSLQSVKVSNTMILIPMKFVIM
FT (in Ref. 1).
SQ SEQUENCE 1455 AA; 162966 MW; 3CB455A99A4FEC86 CRC64;

Query Match 100.0%; Score 104; DB 1; Length 1455;
Best Local Similarity 100.0%; Pred. No. 1.4e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVADLLQ 21
Db 468 DANFDSIRVDVNDVADLLQ 488

RESULT 4
GTFB_STRMU STANDARD; PRT; 1476 AA.
ID -GTFB_STRMU STANDARD; O69384; O69387; O69390; O69396;
AC P08987; O69384; O69387; O69390; O69396;
DT 01-NOV-1988 (Rel. 09, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-MAY-2005 (Rel. 47, Last annotation update)
DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)
DE (Sucrose 6-glucosyltransferase).
GN Name=gtfb; OrderedLocusNames=SMU.1004;
OS Streptococcus mutans.
OS Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OC NCBI_TaxID=1309;
OX [1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=GS-5;
RC MEDLINE=87308013; PubMed=3040685;
RX Shiroza T., Ueda S., Kuramitsu H.K.;
RA "Sequence analysis of the gtfB gene from Streptococcus mutans.";
RT J. Bacteriol. 169:4263-4270(1987).
RL [2]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=MT4239 / Serotype C, MT4245 / Serotype e, MT4251 / Serotype f,
RC MT4467 / Serotype e, and MT8148 / Serotype C;
RX MEDLINE=98231643; PubMed=9570124;
RA Fujiwara T., Terao Y., Hoshino T., Kawabata S., Ooshima T., Sobue S.,
RC Kimura S., Hamada S.;
RT "Molecular analyses of glucosyltransferase genes among strains of
RL Streptococcus mutans.";
RL FEMS Microbiol. Lett. 161:331-336(1998).
RN [3]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Aidic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lal H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
RT pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
CC -1- FUNCTION: Production of extracellular glucans, that are thought to
CC play a key role in the development of the dental plaque because of

```

```

CC their ability to adhere to smooth surfaces and mediate the
CC aggregation of bacterial cells and food debris.
CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl) (n) = D-
CC fructose + (1,6-alpha-D-glucosyl) (n+1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
CC 1,3-linked glucose and some 1,6-linkages). GTF-S synthesizes both
CC water-soluble glucans (alpha 1,6-linkage). GTF-SI synthesizes both
CC forms of glucans.
CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 70 family.
CC -1- SIMILARITY: Contains 10 cell wall binding repeats.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
CC EMBL; M17361; AAA89588.1; -; Genomic_DNA.
CC EMBL; D88651; BAA26101.1; -; Genomic_DNA.
CC EMBL; D88654; BAA26105.1; -; Genomic_DNA.
CC EMBL; D88657; BAA26109.1; -; Genomic_DNA.
CC EMBL; D88660; BAA26113.1; -; Genomic_DNA.
CC EMBL; D89977; BAA26119.1; -; Genomic_DNA.
CC EMBL; AE014940; AAN58705.1; -; Genomic_DNA.
CC PIR; B33135; B33135.
CC HSSP; P06653; 1H8G.
CC InterPro; IPR002479; Cell wall bd put.
CC InterPro; IPR003318; Glyco hydro_70.
CC Pfam; PF01473; CW binding_I; 4.
CC Pfam; PF02324; Glyco_hydro_70; 1.
CC Complete proteome; Dental caries; Glycosyltransferase; Repeat; Signal;
KW Transferase.
KW SIGNAL 1 34 Potential.
FT CHAIN 35 1476 Glucosyltransferase-I.
FT REPEAT 1097 1130 A repeat.
FT REPEAT 1161 1210 1.
FT REPEAT 1225 1275 2.
FT REPEAT 1290 1340 3.
FT REPEAT 1355 1405 4.
FT REPEAT 1420 1470 5.
FT REGION 35 1051 Catalytic (approximate).
FT REGION 1097 1476 Glucan-binding (approximate).
FT REGION 1161 1470 S X tandem repeats.
FT VARIANT 62 62 S -> T (in strain MT4239).
FT VARIANT 65 65 T -> I (in strain GS-5).
FT VARIANT 68 68 V -> A (in strain GS-5, strain MT4245,
FT strain MT4251, strain MT4467 and strain
FT MT8148).
FT VARIANT 78 78 Q -> P (in strain MT4251).
FT VARIANT 86 86 I -> S (in strain GS-5, strain MT4245,
FT strain MT4251, strain MT4467 and strain
FT MT8148).
FT VARIANT 89 89 S -> F (in strain MT4251).
FT VARIANT 168 168 K -> N (in strain MT4251).
FT VARIANT 276 276 S -> D (in strain GS-5, strain MT4467 and
FT strain MT8148).
FT VARIANT 399 399 N -> R (in strain MT4239).
FT VARIANT 474 474 I -> T (in strain MT4239).
FT VARIANT 512 512 K -> R (in strain MT8148).
FT VARIANT 519 519 F -> Y (in strain MT8148).
FT VARIANT 701 701 T -> I (in strain MT8148).
FT VARIANT 708 708 A -> V (in strain MT8148).
FT VARIANT 938 938 F -> L (in strain MT8148).
FT VARIANT 952 957 FGKPEV -> VGTPEA (in strain GS-5, strain
FT MT4239 and strain MT4467).
FT VARIANT 963 964 SV -> NT (in strain GS-5, strain MT4239
FT and strain MT4467).
FT VARIANT 968 970 ADS -> VDG (in strain GS-5, strain MT4239
FT and strain MT4467).
FT VARIANT 1086 1086 A -> T (in strain MT4239).
FT VARIANT 1158 1158 S -> N (in strain MT4239).
FT VARIANT 1163 1163 H -> Y (in strain MT4251).

```

FT VARIANT 1168 1168 E -> K (in strain MT8148).  
 FT VARIANT 1182 1182 Y -> C (in strain MT8148).  
 FT VARIANT 1234 1234 A -> P (in strain MT4239).  
 FT VARIANT 1263 1263 R -> H (in strain GS-5 and strain MT4467).  
 FT VARIANT 1263 1263 R -> P (in strain MT8148).  
 FT VARIANT 1264 1264 Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).  
 FT VARIANT 1272 1272 S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).  
 FT VARIANT 1329 1329 H -> Y (in strain GS-5 and strain MT4467).  
 FT VARIANT 1394 1394 Y -> H (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).  
 FT VARIANT 1402 1402 S -> G (in strain GS-5, strain MT4239, strain MT4467 and strain MT8148).  
 FT VARIANT 1459 1459 Y -> H (in strain MT4467).  
 FT CONFLICT 570 570 R -> A (in Ref. 1).  
 FT CONFLICT 800 817 ADDVRVAASTAPSTDGK -> LKMFALRLARPHQOMA (in Ref. 1).  
 FT CONFLICT 1310 1310 H -> L (in Ref. 1).  
 SQ SEQUENCE 1476 AA; 165847 MW; 9C6E09F731B4CBCF CRC64;

Query Match 100.0%; Score 104; DB 1; Length 1476;  
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQ 21  
 DB 442 DANFDSIRVDVNDVADLLQ 462

RESULT 5  
 Q5263\_9STRE PRELIMINARY; PRT; 1590 AA.  
 AC Q5263;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-JAN-1998 (T-EMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)  
 DE GTF-1.  
 GN Names:Glucosyltransferase;  
 OS Streptococcus sobrinus.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1310;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=ATCC 33478;  
 RA Sato S.;  
 RT "DNA and amino-acid sequences of water-insoluble-glucan synthetase produced from Streptococcus sobrinus ATCC 33478.";  
 RL Kagoshima Daigaku Shigakubu Kiyo 16:23-29(1996).  
 RN [2]  
 RP PROTEIN SEQUENCE.  
 RX MEDLINE=91224988; PubMed=1827439;  
 RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;  
 RT "Isolation and sequence of an active-site peptide containing a catalytic aspartic acid from two Streptococcus sobrinus alpha-glucosyltransferases.";  
 RL J. Biol. Chem. 266:8916-8922(1991).  
 DR EMBL; D63570; BAA09792.1; -; Genomic\_DNA.  
 DR PIR; A39841; A39841.  
 DR HSSP; P06653; 1GVN.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR DR Pfam; PR01473; Glyco\_hydro\_70.  
 DR DR Pfam; PR02324; Glyco\_hydro\_70; 1.  
 SQ SEQUENCE 1590 AA; 176058 MW; 9DF7A3F2C6E4FD43 CRC64;

Query Match 100.0%; Score 104; DB 2; Length 1590;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-05;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQ 21  
 DB 438 DANFDSIRVDVNDVADLLQ 458

RESULT 6  
 Q5983\_9STRE PRELIMINARY; PRT; 1590 AA.  
 ID Q5983;  
 AC Q5983;  
 DT 01-NOV-1996 (T-EMBLrel. 01, Created)  
 DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)  
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)  
 DE Glucosyltransferase-I precursor (EC 2.4.1.5).  
 GN Name=gtfI;  
 OS Streptococcus sobrinus.  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1310;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=OMZ176;  
 RX MEDLINE=94146405; PubMed=8312602;  
 RA Sato S., Inoue M., Handa N., Aizawa Y., Isobe Y., Katayama T.;  
 RT "DNA sequence of the glucosyltransferase gene of serotype d Streptococcus sobrinus.";  
 RL DNA Seq. 4:19-27(1993).  
 RN [2]  
 RP PROTEIN SEQUENCE.  
 RX MEDLINE=91224988; PubMed=1827439;  
 RA Mooser G., Hefta S.A., Paxton R.J., Shively J.E., Lee T.D.;  
 RT "Isolation and sequence of an active-site peptide containing a catalytic aspartic acid from two Streptococcus sobrinus alpha-glucosyltransferases.";  
 RL J. Biol. Chem. 266:8916-8922(1991).  
 DR EMBL; D13858; BAA02976.1; -; Genomic\_DNA.  
 DR PIR; A39841; A39841.  
 DR HSSP; P06653; LHGX.  
 DR GO; GO:0047849; F:dextranucrase activity; IEA.  
 DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.  
 DR GO; GO:0009250; P:glucan biosynthesis; IEA.  
 DR InterPro; IPR002479; CW binding.  
 DR InterPro; IPR003318; Glyco\_hydro\_70.  
 DR DR Pfam; PR01473; CW binding\_1; 2.  
 DR DR Pfam; PR02324; Glyco\_hydro\_70; 1.  
 KW Glucosyltransferase; Signal; Transferase.  
 FT SIGNAL 1 38 Potential.  
 FT CHAIN 39 1590 Glucosyltransferase-I.  
 SQ SEQUENCE 1590 AA; 175956 MW; C3C83A57CF3C2B0E CRC64;

Query Match 100.0%; Score 104; DB 2; Length 1590;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-05;  
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVADLLQ 21  
 DB 438 DANFDSIRVDVNDVADLLQ 458

RESULT 7  
 GTF1\_STRDO STANDARD; PRT; 1597 AA.  
 ID GTF1\_STRDO  
 AC P11001;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DT 13-SEP-2005 (Rel. 48, Last annotation update)  
 DE Glucosyltransferase-I precursor (EC 2.4.1.5) (GTF-I) (Dextranucrase)  
 DE (Sucrose 6-glucosyltransferase).  
 GN Name=gtfI;  
 OS Streptococcus downei (Streptococcus sobrinus).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1317;



```

RN NUCLEOTIDE SEQUENCE [GENOMIC DNA].
RC STRAIN=MFE28;
RA MEDLINE=87308014; PubMed=3040686;
RX Ferretti J.J., Gilpin M.L., Russell R.R.B.;
RT "Nucleotide sequence of a glucosyltransferase gene from Streptococcus
sobrinus MFE28.";
RL Bacteriol. 169:4271-4278(1987).
CC -1- FUNCTION: Production of extracellular glucans, that are thought to
play a key role in the development of the dental plaque because of
their ability to adhere to smooth surfaces and mediate the
aggregation of bacterial cells and food debris.
CC -1- CATALYTIC ACTIVITY: Sucrose + (1,6-alpha-D-glucosyl)(n) = D-
fructose + (1,6-alpha-D-glucosyl)(n+1).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: GTF-I synthesizes water-insoluble glucans (alpha
1,3-linked glucose and some 1,6 linkages), GTF-S synthesizes
water-soluble glucans (alpha 1,6-glucose). GTF-SI synthesizes both
forms of glucans.
CC -1- SIMILARITY: Belongs to the glucosyl hydrolase 70 family.
CC -1- SIMILARITY: Contains 19 cell wall binding repeats.
CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use as long as its content is in no way modified and this statement is not
removed.
CC
CC EMBL; M17391; AAC63063.1; -; Genomic DNA.
DR InterPro; IPR002479; Cell wall bd put.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 4.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Dental caries; Glycosyltransferase; Repeat; Signal; Transferase.
FT SIGNAL 1 38
FT CHAIN 39 1597
FT REPEAT 1099 1132 A repeat.
FT REPEAT 1163 1213 AC repeat.
FT REPEAT 1227 1277 AC repeat.
FT REPEAT 1292 1342 AC repeat.
FT REPEAT 1352 1399 B repeat.
FT REPEAT 1406 1455 AC repeat.
FT REPEAT 1465 1512 B repeat.
FT REPEAT 1519 1568 AC repeat.
FT REPEAT 1582 1597 A repeat (approximate).
FT REGION 39 1050 Catalytic (approximate).
FT REGION 1099 1597 1.25 A, 2 B and 5 AC repeats.
FT REGION 1099 1597 Glucan-binding (approximate).
SQ SEQUENCE 1597 AA; 177080 MW; B9E86A200868798E CRC64;

Query Match 100.0%; Score 104; DB 1; Length 1597;
Best Local Similarity 100.0%; Pred. No. 1.6e-05;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQ 21
DB 444 DANFDSIRVDVNDVNDADLLQ 464

RESULT 8
QSSBM6_LACFE PRELIMINARY; PRT; 1463 AA.
ID QSSBM6;
AC QSSBM6;
DT 01-FEB-2005 (TrEMBLrel. 29, Created)
DT 01-FEB-2005 (TrEMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Glucanucrase [EC 2.4.1.5].
OS Lactobacillus fermentum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1613;
RN [1]
RP NUCLEOTIDE SEQUENCE.

```

```

RC STRAIN=Kg3;
RX PubMed=1528655; DOI=10.1099/mic.0.27321-0;
RA Kralj S., van Geel-Schutten G.H., Dondorff M.M.G., Kirsanovs S.,
van der Maarel M.J.E.C., Dijkhuizen L.;
RT "Glucan synthesis in the genus Lactobacillus: isolation and
characterization of glucanucrase genes, enzymes and glucan products
from six different strains.";
RL Microbiology 150:3681-3690(2004).
DR EMBL; AY697433; AA08008.1; -; Genomic DNA.
DR GO; GO:0047849; P:dextranase activity; IEA.
DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1463 AA; 161049 MW; F7EAB56AB1A3A68 CRC64;

Query Match 96.2%; Score 100; DB 2; Length 1463;
Best Local Similarity 90.5%; Pred. No. 5.2e-05;
Matches 19; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQ 21
DB 563 DANFDSIRVDVNDVNDADLLQ 583

RESULT 9
Q6TXV4_LEUME
ID Q6TXV4_LEUME PRELIMINARY; PRT; 1522 AA.
AC Q6TXV4;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Dextranucrase [EC 2.4.1.5].
GN Name=dsrX;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=L0309;
RA Wang J.-H., Shao Y.-C., Teng D., Yang Y.-L., Zhang F.;
RT "Cloning and sequence analysis of gene coding for dextranucrase from
Leuconostoc mesenteroides L0309.";
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY743959; AAQ98615.2; -; Genomic DNA.
DR GO; GO:0047849; P:dextranase activity; IEA.
DR GO; GO:0016757; P:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW_binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1522 AA; 169103 MW; 01BCC15468B913AE CRC64;

Query Match 96.2%; Score 100; DB 2; Length 1522;
Best Local Similarity 95.2%; Pred. No. 5.4e-05;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQ 21
DB 537 DANFDSIRVDVNDVNDADLLQ 557

RESULT 10
Q8KRE1_LEUME
ID Q8KRE1_LEUME PRELIMINARY; PRT; 1527 AA.
AC Q8KRE1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Dextranucrase Dsrd [EC 2.4.1.5].

```

```

GN Name=dstd;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=22573396; PubMed=12686639; DOI=10.1099/mic.0.26029-0;
RA Neubauer H., Bauche A., Mollet B.;
RT "Molecular characterization and expression analysis of the
RT dextranucrase Dsrd of Leuconostoc mesenteroides Lcc4 in homologous
RT and heterologous Lactococcus lactis cultures.";
RL Microbiology 149:973-982(2003).
DR EMBL; AY017384; AAG61158.1; -; Genomic_DNA.
DR HSSP; P06653; 1H8G.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1527 AA; 169835 MW; F9D0DE220BD89668 CRC64;

Query Match 96.2%; Score 100; DB 2; Length 1527;
Best Local Similarity 95.2%; Pred. No. 5.4e-05;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQ 21
DB 542 DANFDGIRVDVNDVNDADLLQ 562

RESULT 11
Q9ZAR4 LEUME
ID Q9ZAR4 LEUME PRELIMINARY; PRT; 1527 AA.
AC Q9ZAR4;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Dextranucrase.
GN Name=DEX;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX Bhathnagar R., Singh D.K.S.;
RA Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RL EMBL; U81374; AAD10952.1; -; Genomic_DNA.
DR HSSP; P06653; 1H8G.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
SQ SEQUENCE 1527 AA; 169705 MW; 1DFAFA237C743398 CRC64;

Query Match 96.2%; Score 100; DB 2; Length 1527;
Best Local Similarity 95.2%; Pred. No. 5.4e-05;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQ 21
DB 542 DANFDGIRVDVNDVNDADLLQ 562

RESULT 12
Q5SBN0 LACRE
ID Q5SBN0 LACRE PRELIMINARY; PRT; 1772 AA.
AC Q5SBN0;
DT 01-FEB-2005 (T-EMBLrel. 29, Created)

```

```

DT 01-FEB-2005 (T-EMBLrel. 29, Last sequence update)
DT 01-FEB-2005 (T-EMBLrel. 29, Last annotation update)
DE Glucansucrase (EC 2.4.1.5).
OS Lactobacillus reuteri.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1598;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=ML1;
RX PubMed=15528655; DOI=10.1099/mic.0.27321-0;
RA Kralj S., van Geel-Schutten G.H., Dondorff M.M.G., Kirsanovs S.,
RA van der Maarel M.J.E.C., Dijkhuizen L.;
RT "Glucan synthesis in the genus Lactobacillus: isolation and
RT characterization of glucansucrase genes, enzymes and glucan products
RT from six different strains.";
RL Microbiology 150:3681-3690(2004).
DR EMBL; AV697431; AAU08004.1; -; Genomic_DNA.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
DR GO; GO:0016798; F:hydrolase activity, acting on glycosyl bonds; IEA.
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
DR GO; GO:0030245; P:cellulose catabolism; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW_binding_1; 2_70.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1772 AA; 196221 MW; 4A553F1BAD5A210B CRC64;

Query Match 95.2%; Score 99; DB 2; Length 1772;
Best Local Similarity 100.0%; Pred. No. 8.6e-05;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLL 20
DB 1016 DANFDGIRVDVNDVNDADLL 1035

RESULT 13
Q00600 STRSL
ID Q00600 STRSL PRELIMINARY; PRT; 1518 AA.
AC Q00600;
DT 01-NOV-1996 (T-EMBLrel. 01, Created)
DT 01-NOV-1996 (T-EMBLrel. 01, Last sequence update)
DT 01-FEB-2005 (T-EMBLrel. 29, Last annotation update)
DE Glucosyltransferase-I.
GN Name=gtfJ;
OS Streptococcus salivarius.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1304;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=ATCC 25975;
RX MEDLINE=92148377; PubMed=1838391;
RA Giffard P.M., Simpson C.L., Milward C.P., Jacques N.A.;
RT "Molecular characterization of a cluster of at least tow
RT glucosyltransferase genes in Streptococcus salivarius ATCC 25975.";
RL J. Gen. Microbiol. 137:2577-2593(1991).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RX STRAIN=ATCC 25975;
RA Jacques N.;
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z11873; CAA77900.1; -; Genomic_DNA.
DR EMBL; M64111; AAA26896.1; -; Genomic_DNA.
DR FIR; A44811; A44811.
DR HSSP; P06653; 1GVM.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.

```

```
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Transferase.
SQ SEQUENCE 1518 AA; 167730 MW; DAA41F717098B59A CRC64;

Query Match 94.2%; Score 98; DB 2; Length 1518;
Best Local Similarity 90.5%; Pred. No. 0.0001;
Matches 19; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVAVNDVADLLQ 21
Db 462 DANFDGIRVDVAVNDVADMLQ 482

RESULT 14
Q9WXJ5_9STRE
ID Q9WXJ5_9STRE PRELIMINARY; PRT; 1512 AA.
AC Q9WXJ5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GTF-S.
GN Name=gTf;
OS Streptococcus criceti.
OG Plasmid pAM1.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1333;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=HS-6;
RA Inoue M., Fukui K., Miyagi A.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB026123; BAA77237.1; -; Genomic_DNA.
DR HSP; P08654; IMPE.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 2.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Plasmid.
SQ SEQUENCE 1512 AA; 167145 MW; 4C03D9CBC601FC14 CRC64;

Query Match 91.3%; Score 95; DB 2; Length 1512;
Best Local Similarity 90.5%; Pred. No. 0.00026;
Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVAVNDVADLLQ 21
Db 446 DANFDGIRVDVAVNDVADTLQ 466

RESULT 15
Q84CN4_LEUME
ID Q84CN4_LEUME PRELIMINARY; PRT; 1330 AA.
AC Q84CN4;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Dextranucrase DsrR (EC 2.4.1.5).
GN Name=dsr;
OS Leuconostoc mesenteroides.
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.
OX NCBI_TaxID=1245;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=NRRL B-1501;
RA Kim C.H., Moon J.O., Jang E.K.;
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY142210; AAN38835.1; -; Genomic_DNA.
DR GO; GO:0047849; F:dextranucrase activity; IEA.
```

```
DR GO; GO:0016757; F:transferase activity, transferring glycosyl. . .; IEA.
DR GO; GO:0009250; P:glucan biosynthesis; IEA.
DR InterPro; IPR002479; CW binding.
DR InterPro; IPR003318; Glyco_hydro_70.
DR Pfam; PF01473; CW binding_1; 1.
DR Pfam; PF02324; Glyco_hydro_70; 1.
KW Glycosyltransferase; Transferase.
SQ SEQUENCE 1330 AA; 148863 MW; D945CBB36CF75797 CRC64;

Query Match 90.4%; Score 94; DB 2; Length 1330;
Best Local Similarity 95.0%; Pred. No. 0.00032;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ANFDSIRVDVAVNDVADLLQ 21
Db 347 ANFDGIRVDVAVNDVADLLQ 366

Search completed: February 10, 2006, 22:31:11
Job time : 15.4163 secs
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:31:39 ; Search time 3.78931 Seconds  
(without alignments)  
458.180 Million cell updates/sec

Title: US-10-797-821-24

Perfect score: 104

Sequence: 1 DANFDSIRVDVNDVDALLQ 21

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5\_COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/6\_COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/H\_COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/RE\_COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	22	1	US-08-057-162B-1
2	104	100.0	1375	2	US-09-210-361-4
3	104	100.0	1375	2	US-09-740-274-4
4	104	100.0	1475	2	US-09-007-999-2
5	104	100.0	1475	2	US-09-210-361-2
6	104	100.0	1475	2	US-09-740-274-2
7	100	96.2	522	2	US-09-995-749A-11
8	100	96.2	523	2	US-09-604-957-5
9	92	88.5	1577	1	US-08-793-824-2
10	91	87.5	584	2	US-09-604-957-6
11	91	87.5	584	2	US-09-995-749A-12
12	91	87.5	2057	2	US-09-499-203-2
13	90	86.5	545	2	US-09-604-957-4
14	90	86.5	545	2	US-09-995-749A-10
15	90	86.5	1430	2	US-09-008-172-2
16	90	86.5	1430	2	US-09-210-361-6
17	90	86.5	1430	2	US-09-740-274-6
18	83	79.8	535	2	US-09-604-957-7
19	83	79.8	535	2	US-09-995-749A-13
20	83	79.8	1278	2	US-09-604-957-3
21	83	79.8	1781	2	US-09-995-749A-2
22	54	51.9	624	1	US-08-947-965-78
23	54	51.9	655	1	US-08-469-202-27
24	54	51.9	655	1	US-08-469-202-28
25	54	51.9	655	1	US-08-484-434C-34
26	54	51.9	655	1	US-08-484-434C-35
27	54	51.9	655	2	US-09-384-361-34

28 54 51.9 655 2 US-09-384-361-35 Sequence 35, Appli  
29 46 44.2 25 2 US-08-888-080A-6 Sequence 6, Appli  
30 46 44.2 25 2 US-09-577-027-6 Sequence 6, Appli  
31 46 44.2 38 2 US-08-888-080A-7 Sequence 7, Appli  
32 46 44.2 38 2 US-08-888-080A-8 Sequence 8, Appli  
33 46 44.2 38 2 US-09-577-027-7 Sequence 7, Appli  
34 46 44.2 38 2 US-09-577-027-8 Sequence 8, Appli  
35 45 43.3 465 2 US-09-252-991A-21721 Sequence 21721, A  
36 45 43.3 793 2 US-09-489-039A-7449 Sequence 7449, Ap  
37 44 42.3 25 2 US-08-888-080A-4 Sequence 4, Appli  
38 44 42.3 25 2 US-09-577-027-4 Sequence 4, Appli  
39 43.5 41.8 3177 1 US-08-477-451-4 Sequence 4, Appli  
40 43 41.3 330 2 US-09-252-991A-24853 Sequence 24853, A  
41 43 41.3 668 2 US-09-538-092-599 Sequence 599, App  
42 43 41.3 762 2 US-09-786-480B-25 Sequence 25, Appli  
43 42 40.4 25 2 US-08-888-080A-5 Sequence 5, Appli  
44 42 40.4 25 2 US-09-577-027-5 Sequence 5, Appli  
45 42 40.4 29 2 US-08-987-912-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-057-162B-1  
; Sequence 1, Application US/08057162B  
; Patent No. 5686075  
; GENERAL INFORMATION:  
; APPLICANT: Taubman, Martin A.  
; APPLICANT: Smith, Daniel J.  
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR DENTAL CARIES  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/057,162B  
; FILING DATE: 30-APR-1993  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/877,295  
; FILING DATE: 01-MAY-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wagner, Richard W.  
; REGISTRATION NUMBER: 34,480  
; REFERENCE/DOCKET NUMBER: FDC92-01A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 861-6240  
; TELEFAX: (617) 861-9540  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 22 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-057-162B-1

Query Match 100.0%; Score 104; DB 1; Length 22;  
Best Local Similarity 100.0%; Pred. No. 1.3e-10;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVDALLQ 21  
|||||

```
Db      1 DANFDSIRVDAVNDVNDADLLQ 21
RESULT 2
US-09-210-361-4
; Sequence 4, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; EARLIER FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 1375
; TYPE: PRT
; ORGANISM: streptococcus mutans
US-09-210-361-4
Query Match      100.0%; Score 104; DB 2; Length 1375;
Best Local Similarity 100.0%; Pred. No. 2.1e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DANFDSIRVDAVNDVNDADLLQ 21
Db      468 DANFDSIRVDAVNDVNDADLLQ 488
RESULT 4
US-09-007-999-2
; Sequence 2, Application US/09007999
; Patent No. 6087559
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starch and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0356D
; CURRENT APPLICATION NUMBER: US/09/007,999
; CURRENT FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-007-999-2
Query Match      100.0%; Score 104; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DANFDSIRVDAVNDVNDADLLQ 21
Db      442 DANFDSIRVDAVNDVNDADLLQ 462
RESULT 5
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2
Query Match      100.0%; Score 104; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DANFDSIRVDAVNDVNDADLLQ 21
Db      442 DANFDSIRVDAVNDVNDADLLQ 462
RESULT 5
US-09-210-361-2
; Sequence 2, Application US/09210361
; Patent No. 6284479
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Substitutes for Modified Starches and
; TITLE OF INVENTION: Latexes in Paper Manufacture
; FILE REFERENCE: 0357CR
; CURRENT APPLICATION NUMBER: US/09/210,361
; CURRENT FILING DATE: 1998-12-11
; EARLIER APPLICATION NUMBER: 09/007,999
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/478,704
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/009,620
; EARLIER FILING DATE: 1998-01-20
; EARLIER APPLICATION NUMBER: 08/485,243
; EARLIER FILING DATE: 1995-06-07
; EARLIER APPLICATION NUMBER: 09/008,172
; EARLIER FILING DATE: 1998-01-16
; EARLIER APPLICATION NUMBER: 08/482,711
; EARLIER FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-210-361-2
Query Match      100.0%; Score 104; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVDADLLQ 21
Db 442 DANFDSIRVDVNDVDADLLQ 462

RESULT 6
US-09-740-274-2
; Sequence 2, Application US/09740274
; Patent No. 6465203
; GENERAL INFORMATION:
; APPLICANT: Nichols, Scott E.
; TITLE OF INVENTION: Glucan-containing Compositions and Paper
; FILE REFERENCE: 0357CRD
; CURRENT APPLICATION NUMBER: US/09/740,274
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 09/210,361
; PRIOR FILING DATE: 1998-12-11
; PRIOR APPLICATION NUMBER: 09/007,999
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/478,704
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/009,620
; PRIOR FILING DATE: 1998-01-20
; PRIOR APPLICATION NUMBER: 08/485,243
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: 09/008,172
; PRIOR FILING DATE: 1998-01-16
; PRIOR APPLICATION NUMBER: 08/482,711
; PRIOR FILING DATE: 1995-06-07
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-740-274-2

Query Match 100.0%; Score 104; DB 2; Length 1475;
Best Local Similarity 100.0%; Pred. No. 2.3e-08;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVDADLLQ 21
Db 442 DANFDSIRVDVNDVDADLLQ 462

RESULT 7
US-09-995-749A-11
; Sequence 11, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749A-11

Best Local Similarity 96.2%; Score 100; DB 2; Length 522;
Best Local Similarity 95.2%; Pred. No. 2.9e-08;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVDADLLQ 21
Db 36 DANFDGIRVDVNDVDADLLQ 56

RESULT 8
US-09-604-957-5
; Sequence 5, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: B0 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 5
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-604-957-5

Query Match 96.2%; Score 100; DB 2; Length 523;
Best Local Similarity 95.2%; Pred. No. 2.9e-08;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVDADLLQ 21
Db 36 DANFDGIRVDVNDVDADLLQ 56

RESULT 9
US-08-793-824-2
; Sequence 2, Application US/08793824
; Patent No. 5981838
; GENERAL INFORMATION:
; APPLICANT: Simpson, Christine Lynn
; APPLICANT: Giffard, Philip Morrison
; APPLICANT: Jacques, Nicholas Anthony
; TITLE OF INVENTION: Genetic Manipulation of Plants to
; TITLE OF INVENTION: Increase Stored Carbohydrates
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Griffith Hack & Co
; STREET: Level 8, 168 Walker Street
; CITY: No. 5981838th Sydney
; STATE: New South Wales
; COUNTRY: Australia
; ZIP: 2060
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,824
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: AU PM7643
; FILING DATE: 24-AUG-1994
; TELECOMMUNICATION INFORMATION:
```

```
; TELEPHONE: 61 2 9957 5944
; TELEFAX: 61 2 957 6288
; TELEX: 26547
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1577 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus salivarius
US-08-793-824-2

Query Match      88.5%; Score 92; DB 1; Length 1577;
Best Local Similarity 85.0%; Pred. No. 2.4e-06;
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      2 ANFDSIRVDVNDVNDADLLQ 21
Db      550 ANFDGVRIDAVNDVNDADLLQ 569

RESULT 10
US-09-604-957-6
; Sequence 6, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-604-957-6

Query Match      87.5%; Score 91; DB 2; Length 584;
Best Local Similarity 85.7%; Pred. No. 1e-06;
Matches 18; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 DANFDSIRVDVNDVNDADLLQ 21
Db      36 NANFDGIRVDVNDVNDADLLK 56

RESULT 11
US-09-995-749A-12
; Sequence 12, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: BO43388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: EPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19

Query Match      88.5%; Score 92; DB 1; Length 1577;
Best Local Similarity 85.0%; Pred. No. 2.4e-06;
Matches 17; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      2 ANFDSIRVDVNDVNDADLLQ 21
Db      550 ANFDGVRIDAVNDVNDADLLQ 569

RESULT 10
US-09-604-957-6
; Sequence 6, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-604-957-6

Query Match      87.5%; Score 91; DB 2; Length 584;
Best Local Similarity 85.7%; Pred. No. 1e-06;
Matches 18; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 DANFDSIRVDVNDVNDADLLQ 21
Db      36 NANFDGIRVDVNDVNDADLLK 56

RESULT 11
US-09-995-749A-12
; Sequence 12, Application US/09995749A
; Patent No. 6867026
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19

Query Match      86.5%; Score 90; DB 2; Length 545;
Best Local Similarity 81.0%; Pred. No. 1.4e-06;
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 584
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749A-12

Query Match      87.5%; Score 91; DB 2; Length 584;
Best Local Similarity 85.7%; Pred. No. 1e-06;
Matches 18; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 DANFDSIRVDVNDVNDADLLQ 21
Db      36 NANFDGIRVDVNDVNDADLLK 56

RESULT 12
US-09-499-203-2
; Sequence 2, Application US/09499203
; Patent No. 6570065
; GENERAL INFORMATION:
; APPLICANT: KOSSMANN, Jens
; APPLICANT: WELSH, Thomas
; APPLICANT: QUANZ, Martin
; APPLICANT: KNUTH, Karola
; TITLE OF INVENTION: Nucleic Acid Molecules Encoding Alternansucrase
; FILE REFERENCE: 147-196P
; CURRENT APPLICATION NUMBER: US/09/499,203
; CURRENT FILING DATE: 2000-02-08
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2057
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-499-203-2

Query Match      87.5%; Score 91; DB 2; Length 2057;
Best Local Similarity 85.7%; Pred. No. 4.8e-06;
Matches 18; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY      1 DANFDSIRVDVNDVNDADLLQ 21
Db      626 NANFDGIRVDVNDVNDADLLK 646

RESULT 13
US-09-604-957-4
; Sequence 4, Application US/09604957
; Patent No. 6486314
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAOUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASE AND NOVEL GLUCAN
; FILE REFERENCE: BO 43388
; CURRENT APPLICATION NUMBER: US/09/604,957
; CURRENT FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 545
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-604-957-4

Query Match      86.5%; Score 90; DB 2; Length 545;
Best Local Similarity 81.0%; Pred. No. 1.4e-06;
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
```



Qy 1 DANFDSIRVDVNDVNDADLLQ 21  
:||||:|||||:|||||  
Db 36 EANFDGVRVDVNDVNDADLLQ 56

## RESULT 14

US-09-995-749A-10  
; Sequence 10, Application US/09995749A  
; Patent No. 6867026  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; APPLICANT: DIJKHUIZEN, LUBBERT  
; APPLICANT: RAHAOUI, HAKIM  
; APPLICANT: LEER, ROBERT-JAN  
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES  
; FILE REFERENCE: B043388-CIP  
; CURRENT APPLICATION NUMBER: US/09/995,749A  
; CURRENT FILING DATE: 2001-11-29  
; PRIOR APPLICATION NUMBER: 09/604,957  
; PRIOR FILING DATE: 2000-06-28  
; PRIOR APPLICATION NUMBER: EPO 00201871.1  
; PRIOR FILING DATE: 2000-05-25  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 545  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-995-749A-10

Query Match 86.5%; Score 90; DB 2; Length 545;  
Best Local Similarity 81.0%; Pred. No. 1.4e-06;  
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVNDADLLQ 21  
:||||:|||||:|||||  
Db 36 EANFDGVRVDVNDVNDADLLQ 56

## RESULT 15

US-09-008-172-2  
; Sequence 2, Application US/09008172  
; Patent No. 6127602  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Substitutes for Modified Starch and  
; TITLE OF INVENTION: Latexes in Paper Manufacture  
; FILE REFERENCE: 0358D  
; CURRENT APPLICATION NUMBER: US/09/008,172  
; CURRENT FILING DATE: 1998-01-16  
; EARLIER APPLICATION NUMBER: 08/482,711  
; EARLIER FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1430  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-008-172-2

Query Match 86.5%; Score 90; DB 2; Length 1430;  
Best Local Similarity 81.0%; Pred. No. 4.5e-06;  
Matches 17; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVNDADLLQ 21  
:||||:|||||:|||||  
Db 456 EANFDGVRVDVNDVNDADLLQ 476

Search completed: February 10, 2006, 22:36:13  
Job time : 3.78931 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:14:04 ; Search time 12.3206 Seconds  
(without alignments)  
712.176 Million cell updates/sec

Title: US-10-797-821-24  
Perfect score: 104  
Sequence: 1 DANFDSIRVDVNDVADLLQ 21

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA Main.\*  
1: /cgm2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgm2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
3: /cgm2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
4: /cgm2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
5: /cgm2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
6: /cgm2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	100.0	21	4	US-10-383-930-24
2	104	100.0	21	5	US-10-797-821-24
3	104	100.0	22	4	US-10-383-930-25
4	104	100.0	22	5	US-10-797-821-25
5	104	100.0	1375	3	US-09-740-274-4
6	104	100.0	1375	4	US-10-383-930-35
7	104	100.0	1375	5	US-10-797-821-35
8	104	100.0	1475	3	US-09-740-274-2
9	104	100.0	1475	4	US-10-383-930-34
10	104	100.0	1475	5	US-10-797-821-34
11	104	100.0	1590	4	US-10-383-930-37
12	104	100.0	1590	5	US-10-797-821-37
13	100	96.2	522	3	US-09-995-749A-11
14	100	96.2	1006	5	US-10-484-218-22
15	99	95.2	221	5	US-10-484-218-8
16	99	95.2	1771	5	US-10-484-218-14
17	98	94.2	23	2	US-08-967-573A-11
18	98	94.2	1518	4	US-10-383-930-40
19	98	94.2	1518	5	US-10-797-821-40
20	94	90.4	223	5	US-10-484-218-6
21	94	90.4	1365	4	US-10-383-930-39
22	94	90.4	1365	5	US-10-797-821-39
23	94	90.4	1554	4	US-10-383-930-38
24	94	90.4	1554	5	US-10-797-821-38
25	92	88.5	1595	5	US-10-484-218-20
26	91	87.5	584	3	US-09-995-749A-12
27	91	87.5	2057	4	US-10-417-280A-2

28 90 86.5 545 3 US-09-995-749A-10 Sequence 10, Appli  
29 90 86.5 1430 3 US-09-740-274-6 Sequence 6, Appli  
30 90 86.5 1430 4 US-10-383-930-36 Sequence 36, Appli  
31 90 86.5 1430 5 US-10-797-821-36 Sequence 36, Appli  
32 89 85.6 223 5 US-10-484-218-10 Sequence 10, Appli  
33 87 83.7 221 5 US-10-484-218-2 Sequence 2, Appli  
34 87 83.7 525 5 US-10-484-218-23 Sequence 23, Appli  
35 87 83.7 1777 5 US-10-484-218-12 Sequence 12, Appli  
36 84 80.8 224 5 US-10-484-218-4 Sequence 4, Appli  
37 84 80.8 1497 5 US-10-484-218-18 Sequence 18, Appli  
38 83 79.8 535 3 US-09-995-749A-13 Sequence 13, Appli  
39 83 79.8 1781 3 US-09-995-749A-2 Sequence 2, Appli  
40 79 76.0 787 5 US-10-484-218-16 Sequence 16, Appli  
41 50 48.1 590 4 US-10-425-115-310890 Sequence 310890,  
42 46 44.2 25 3 US-09-922-568-6 Sequence 6, Appli  
43 46 44.2 38 3 US-09-922-568-7 Sequence 7, Appli  
44 46 44.2 38 3 US-09-922-568-8 Sequence 8, Appli  
45 46 44.2 350 4 US-10-437-963-119366 Sequence 119366,

ALIGNMENTS

RESULT 1  
US-10-383-930-24  
; Sequence 24, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 24  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-24

Query Match 100.0%; Score 104; DB 4; Length 21;  
Best Local Similarity 100.0%; Pred. No. 5.3e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVADLLQ 21  
| | | | | | | | | | | | | | | | | | | | |  
Db 1 DANFDSIRVDVNDVADLLQ 21

RESULT 2  
US-10-797-821-24  
; Sequence 24, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049

; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 24  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-24

Query Match 100.0%; Score 104; DB 5; Length 21;  
Best Local Similarity 100.0%; Pred. No. 5.3e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQ 21  
|||||  
Db 1 DANFDSIRVDVNDVNDADLLQ 21

## RESULT 3

US-10-383-930-25  
; Sequence 25, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 25  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-25

Query Match 100.0%; Score 104; DB 4; Length 22;  
Best Local Similarity 100.0%; Pred. No. 5.6e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQ 21  
|||||  
Db 1 DANFDSIRVDVNDVNDADLLQ 21

## RESULT 4

US-10-797-821-25  
; Sequence 25, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12

; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 25  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: GTF-derived catalytic (CAT) peptide  
US-10-797-821-25

Query Match 100.0%; Score 104; DB 5; Length 22;  
Best Local Similarity 100.0%; Pred. No. 5.6e-09;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQ 21  
|||||  
Db 1 DANFDSIRVDVNDVNDADLLQ 21

## RESULT 5

US-09-740-274-4  
; Sequence 4, Application US/09740274  
; Patent No. US20020031826A1  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19  
; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: streptococcus mutans  
US-09-740-274-4

Query Match 100.0%; Score 104; DB 3; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 6.7e-07;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DANFDSIRVDVNDVNDADLLQ 21  
|||||  
Db 468 DANFDSIRVDVNDVNDADLLQ 488

## RESULT 6

US-10-383-930-35  
; Sequence 35, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018

; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 35  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-35

Query Match 100.0%; Score 104; DB 4; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 6.7e-07;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVADLLQ 21  
Db 468 DANFDSIRVDVNDVADLLQ 488  
|||||

## RESULT 7

US-10-797-821-35  
; Sequence 35, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049  
; PRIOR FILING DATE: 1999-04-12  
; PRIOR APPLICATION NUMBER: 60/081,550  
; PRIOR FILING DATE: 1998-04-13  
; PRIOR APPLICATION NUMBER: 60/115,142  
; PRIOR FILING DATE: 1999-01-08  
; NUMBER OF SEQ ID NOS: 45  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 35  
; LENGTH: 1375  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-797-821-35

Query Match 100.0%; Score 104; DB 5; Length 1375;  
Best Local Similarity 100.0%; Pred. No. 6.7e-07;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVADLLQ 21  
Db 468 DANFDSIRVDVNDVADLLQ 488  
|||||

## RESULT 8

US-09-740-274-2  
; Sequence 2, Application US/09740274  
; Patent No. US20020031826A1  
; GENERAL INFORMATION:  
; APPLICANT: Nichols, Scott E.  
; TITLE OF INVENTION: Glucan-containing Compositions and Paper  
; FILE REFERENCE: 0357CRD  
; CURRENT APPLICATION NUMBER: US/09/740,274  
; CURRENT FILING DATE: 2000-12-19

; PRIOR APPLICATION NUMBER: 09/210,361  
; PRIOR FILING DATE: 1998-12-11  
; PRIOR APPLICATION NUMBER: 09/007,999  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/478,704  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/009,620  
; PRIOR FILING DATE: 1998-01-20  
; PRIOR APPLICATION NUMBER: 08/485,243  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: 09/008,172  
; PRIOR FILING DATE: 1998-01-16  
; PRIOR APPLICATION NUMBER: 08/482,711  
; PRIOR FILING DATE: 1995-06-07  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-09-740-274-2

Query Match 100.0%; Score 104; DB 3; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 7.2e-07;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVADLLQ 21  
Db 442 DANFDSIRVDVNDVADLLQ 462  
|||||

## RESULT 9

US-10-383-930-34  
; Sequence 34, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 34  
; LENGTH: 1475  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
US-10-383-930-34

Query Match 100.0%; Score 104; DB 4; Length 1475;  
Best Local Similarity 100.0%; Pred. No. 7.2e-07;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVADLLQ 21  
Db 442 DANFDSIRVDVNDVADLLQ 462  
|||||

## RESULT 10

US-10-797-821-34  
; Sequence 34, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821

```
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 1475
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-34

Query Match          100.0%; Score 104; DB 5; Length 1475;
Best Local Similarity 100.0%; Pred. No. 7.2e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVNDADLLQ 21
   ||||| ||||| ||||| ||||| |||||
Db 442 DANFDSIRVDVNDVNDADLLQ 462

RESULT 11
US-10-383-930-37
; Sequence 37, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-383-930-37

Query Match          100.0%; Score 104; DB 4; Length 1590;
Best Local Similarity 100.0%; Pred. No. 7.9e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVNDADLLQ 21
   ||||| ||||| ||||| ||||| |||||
Db 438 DANFDSIRVDVNDVNDADLLQ 458

RESULT 12
US-10-797-821-37
; Sequence 37, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09

; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 37
; LENGTH: 1590
; TYPE: PRT
; ORGANISM: Streptococcus sobrinus
US-10-797-821-37

Query Match          100.0%; Score 104; DB 5; Length 1590;
Best Local Similarity 100.0%; Pred. No. 7.9e-07;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVNDADLLQ 21
   ||||| ||||| ||||| ||||| |||||
Db 438 DANFDSIRVDVNDVNDADLLQ 458

RESULT 13
US-09-995-749A-11
; Sequence 11, Application US/09995749A
; Patent No. US20020155568A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITINA HENDRIKA
; APPLICANT: DIJKHUIZEN, LUBBERT
; APPLICANT: RAHAQUI, HAKIM
; APPLICANT: LEER, ROBERT-JAN
; TITLE OF INVENTION: NOVEL GLUCOSYLTRANSFERASES
; FILE REFERENCE: B043388-CIP
; CURRENT APPLICATION NUMBER: US/09/995,749A
; CURRENT FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 09/604,957
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: BPO 00201871.1
; PRIOR FILING DATE: 2000-05-25
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 522
; TYPE: PRT
; ORGANISM: Leuconostoc mesenteroides
US-09-995-749A-11

Query Match          96.2%; Score 100; DB 3; Length 522;
Best Local Similarity 95.2%; Pred. No. 8.9e-07;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVNDADLLQ 21
   ||||| ||||| ||||| ||||| |||||
Db 36 DANFDSIRVDVNDVNDADLLQ 56

RESULT 14
US-10-484-218-22
; Sequence 22, Application US/10484218
; Publication No. US20050059633A1
; GENERAL INFORMATION:
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITINA HENDRIKA
; TITLE OF INVENTION: GLUCANS AND GLUCANSUCRASES DERIVED FROM
; TITLE OF INVENTION: LACTIC ACID BACTERIA
; FILE REFERENCE: 2001-1316
; CURRENT APPLICATION NUMBER: US/10/484,218
```

; CURRENT FILING DATE: 2004-01-20  
; PRIOR APPLICATION NUMBER: PCT/NL02/00495  
; PRIOR FILING DATE: 2002-07-22  
; PRIOR APPLICATION NUMBER: EP 01202752.0  
; PRIOR FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: EP 01202841.1  
; PRIOR FILING DATE: 2001-07-25  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 22  
; LENGTH: 1006  
; TYPE: PRT  
; ORGANISM: Lactobacillus fermentum  
US-10-484-218-22

Query Match 96.2%; Score 100; DB 5; Length 1006;  
Best Local Similarity 90.5%; Pred.No. 1.9e-06;  
Matches 19; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVDADLLQ 21  
Db 106 DANFDAIRIDAVDNDVDADLLQ 126

RESULT 15  
US-10-484-218-8  
; Sequence 8, Application US/10484218  
; Publication No. US20050059633A1  
; GENERAL INFORMATION:  
; APPLICANT: VAN GEEL-SCHUTTEN, GERRITDINA HENDRIKA  
; TITLE OF INVENTION: GLUCANS AND GLUCANSUCRASES DERIVED FROM  
; TITLE OF INVENTION: LACTIC ACID BACTERIA  
; FILE REFERENCE: 2001-1316  
; CURRENT APPLICATION NUMBER: US/10/484,218  
; CURRENT FILING DATE: 2004-01-20  
; PRIOR APPLICATION NUMBER: PCT/NL02/00495  
; PRIOR FILING DATE: 2002-07-22  
; PRIOR APPLICATION NUMBER: EP 01202752.0  
; PRIOR FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: EP 01202841.1  
; PRIOR FILING DATE: 2001-07-25  
; NUMBER OF SEQ ID NOS: 30  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 8  
; LENGTH: 221  
; TYPE: PRT  
; ORGANISM: Lactobacillus sp.  
US-10-484-218-8

Query Match 95.2%; Score 99; DB 5; Length 221;  
Best Local Similarity 100.0%; Pred.No. 4.7e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DANFDSIRVDVNDVDADLL 20  
Db 30 DANFDSIRVDVNDVDADLL 49

Search completed: February 10, 2006, 23:25:10  
Job time : 13.3206 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:05:33 ; Search time 14.4758 Seconds  
(without alignments)  
607.053 Million cell updates/sec

Title: US-10-797-821-21

Perfect score: 108

Sequence: 1 DGGYGHVAYVTGGGQIQV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A Geneseq\_21.\*

- 1: Geneseqp1980s.\*
- 2: Geneseqp1990s.\*
- 3: Geneseqp2000s.\*
- 4: Geneseqp2001s.\*
- 5: Geneseqp2002s.\*
- 6: Geneseqp2003as.\*
- 7: Geneseqp2003bs.\*
- 8: Geneseqp2004s.\*
- 9: Geneseqp2005s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	100.0	20	7	ADD93641 Streptoco
2	108	100.0	20	9	ADX37264 Streptoco
3	108	100.0	431	7	ADD93649 Streptoco
4	108	100.0	431	7	ADD93650 Streptoco
5	108	100.0	431	7	ADD93653 Streptoco
6	108	100.0	431	9	ADX37272 Streptoco
7	108	100.0	431	9	ADX37273 Streptoco
8	108	100.0	431	9	ADX37276 Streptoco
9	108	100.0	431	9	AE91500 Microbial
10	108	100.0	432	7	ADD93651 Streptoco
11	108	100.0	432	7	ADD93652 Streptoco
12	108	100.0	432	9	ADX37274 Streptoco
13	108	100.0	432	9	ADX37275 Streptoco
14	76.5	70.8	211	9	AE91642 Microbial
15	70.5	65.3	544	9	AE91648 Microbial
16	63	58.3	178	3	AA71288 S. pneumo
17	63	58.3	448	3	AA71285 Streptoco
18	63	58.3	448	6	ABU02732 S. pneumo
19	63	58.3	448	8	ADM92272 S. pneumo
20	63	58.3	448	8	ADT50163 S. pneumo
21	63	58.3	448	9	AE91520 Microbial
22	62.5	57.9	132	2	AAW60944 Streptoco
23	62.5	57.9	392	6	ABU02747 S. pneumo
24	62.5	57.9	392	8	ADK47859 Streptoco

ALIGNMENTS

RESULT 1

ADD93641

ID ADD93641 standard; peptide; 20 AA.

AC ADD93641;

XX

DT 29-JAN-2004 (first entry)

XX

DE Streptococcus mutans glucan binding protein-B peptide fragment.

XX

KW Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX

OS Streptococcus mutans.

XX

PN WC2003075845-A2.

XX

PD 18-SEP-2003.

XX

PF 07-MAR-2003; 2003MO-US006962.

XX

PR 07-MAR-2002; 2002US-0363209P.

XX

PR 08-AUG-2002; 2002US-0402483P.

XX

XX (FORS-) FORSYTH INST.

XX

XX Smith DJ, Taubman MA;

XX

DR WPI; 2003-845091/78.

XX

PT Composition useful as vaccines for dental caries comprises a fragment of a glucan binding protein-B binding to a major histocompatibility complex class II protein.

XX

PS Claim 11; Page 10; 49pp; English.

XX

XX The present sequence is that of a peptide comprising amino acid residues 383-402 of the glucan binding protein-B (GbpB) of Streptococcus mutans strain SK32 AD93649. The peptide binds to a major histocompatibility complex (MHC) class II protein. It was identified as a potential B cell epitope using a matrix-based algorithm for epitope prediction, which was used to search the primary amino acid sequence of GbpB for known MHC class II binding motifs. The peptide can be used in immunogenic compositions and subunit vaccines for dental caries. These comprise MHC class II protein-binding GbpB peptides covalently linked with peptide subunits (preferably from the catalytic domain) of a glucosyltransferase. The compositions are used in a claimed method of eliciting production of an antibody in a mammal. Diepitopic or multiepitopic polypeptides can be

CC prepared synthetically or by recombinant DNA technology. Antibodies  
 CC raised against MHC class II binding fragments of GbpB can be used in  
 CC passive immunisation.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 108; DB 7; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DGGYGHVAVYTVGGGQIQV 20  
 |||||  
 Db 1 DGGYGHVAVYTVGGGQIQV 20

## RESULT 2

ADX37264  
 ID ADX37264 standard; peptide; 20 AA.

XX AC ADX37264;

DT 21-APR-2005 (first entry)

XX Streptococcus mutant glucan binding protein B peptide #21.

DE immunogenicity; immune stimulation; glucan binding protein-B;  
 KW microparticle; major histocompatibility complex; tooth disease.

XX Streptococcus mutans.

OS US2005031633-A1.

XX PN 10-FEB-2005.

XX PF 09-MAR-2004; 2004US-00797821.

XX PR 13-APR-1998; 98US-0081550P.

XX PR 08-JAN-1999; 99US-0115142P.

XX PR 12-APR-1999; 99US-00290049.

XX PR 07-MAR-2002; 2002US-0363209P.

XX PR 08-AUG-2002; 2002US-0402483P.

XX PR 07-MAR-2003; 2003US-00383930.

XX PA (SMIT/J) SMITH D J.

XX PA (TAUB/) TAUBMAN M A.

XX PI Smith DJ, Taubman MA;

XX DR WPI; 2005-151644/16.

XX PT New composition comprising a fragment of a glucan binding protein-B

PT (GbpB) that binds to MHC class II protein, and a biocompatible  
 PT microparticle, useful for producing an antibody (claimed) for immunizing  
 PT mammals against dental caries.

XX PS Claim 4; SEQ ID NO 21; 73pp; English.

XX CC The invention relates to a composition comprising a fragment of a glucan  
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II  
 CC protein. The composition is useful for producing an antibody for  
 CC immunizing mammals against dental caries. This sequence corresponds to a  
 CC Streptococcus mutans GbpB-derived peptide of the invention.

XX SQ Sequence 20 AA;

Query Match 100.0%; Score 108; DB 9; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 5e-09;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 DGGYGHVAVYTVGGGQIQV 20  
 |||||  
 Db 1 DGGYGHVAVYTVGGGQIQV 20

## RESULT 3

ADD93649  
 ID ADD93649 standard; protein; 431 AA.

XX AC ADD93649;

DT 29-JAN-2004 (first entry)

XX Streptococcus mutans glucan binding protein-B.

DE Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX Streptococcus mutans.

XX FH Key Location/Qualifiers

FT Region 6..25  
 /note= "HLA-binding peptide"

FT Region 16..35  
 /note= "HLA-binding peptide"

FT Region 33..52  
 /note= "HLA-binding peptide"

FT Region 37..56  
 /note= "HLA-binding peptide"

FT Region 48..67  
 /note= "HLA-binding peptide"

FT Region 52..71  
 /note= "HLA-binding peptide"

FT Region 88..107  
 /note= "HLA-binding peptide"

FT Region 113..132  
 /note= "HLA-binding peptide"

FT Region 117..136  
 /note= "HLA-binding peptide"

FT Region 137..156  
 /note= "HLA-binding peptide"

FT Region 174..193  
 /note= "HLA-binding peptide"

FT Region 194..213  
 /note= "HLA-binding peptide"

FT Region 214..233  
 /note= "HLA-binding peptide"

FT Region 248..267  
 /note= "HLA-binding peptide"

FT Region 289..308  
 /note= "HLA-binding peptide"

FT Region 306..325  
 /note= "HLA-binding peptide"

FT Region 311..330  
 /note= "HLA-binding peptide"

FT Region 349..368  
 /note= "HLA-binding peptide"

FT Region 365..384  
 /note= "HLA-binding peptide"

FT Region 383..402  
 /note= "HLA-binding peptide"

FT Region 403..422  
 /note= "HLA-binding peptide"

XX WO2003075845-A2.

PN 18-SEP-2003.

XX 07-MAR-2003; 2003WO-US006962.

XX 07-MAR-2002; 2002US-0363209P.

PR 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman MA;

XX

DR WPI; 2003-845091/78.  
 XX GENBANK; AY046410.  
 CC Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.  
 XX Claim 6; Page 7; 49pp; English.  
 XX  
 CC The present sequence is the protein sequence of the glucan binding  
 CC protein-B (GbpB) of Streptococcus mutans strain SK32. The sequence  
 CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
 CC invention provides immunogenic compositions and vaccines for dental  
 CC caries. The compositions comprise major histocompatibility complex (MHC)  
 CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
 CC covalently linked with peptide subunits of a glucosyltransferase. The  
 CC compositions are used in a claimed method of eliciting production of an  
 CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be  
 CC prepared synthetically or by recombinant DNA technology. Antibodies  
 CC raised against MHC class II binding fragments of GbpB can be used in  
 CC passive immunisation.  
 XX Sequence 431 AA;  
 SQ  
 Query Match 100.0%; Score 108; DB 7; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGGYGHVAVYTVGGGQIQV 20  
 DB 383 DGGYGHVAVYTVGGGQIQV 402  
 |||||  
 RESULT 4  
 ADD93650  
 ID ADD93650 standard; protein; 431 AA.  
 XX  
 AC ADD93650;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Streptococcus mutans glucan binding protein-B.  
 XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
 XX Streptococcus mutans.  
 XX WO2003075845-A2.  
 XX PD 18-SEP-2003.  
 XX PF 07-MAR-2003; 2003WO-US006962.  
 XX PR 07-MAR-2002; 2002US-0363209P.  
 XX PR 08-AUG-2002; 2002US-0402483P.  
 XX (FORS-) FORSYTH INST.  
 XX Smith DJ, Taubman MA;  
 XX WPI; 2003-845091/78.  
 XX GENBANK; AY046411.  
 XX Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.  
 XX Claim 5; Page 8; 49pp; English.  
 XX  
 CC The present sequence is the protein sequence of the glucan binding  
 CC protein-B (GbpB) of Streptococcus mutans strain 3VF4. The sequence  
 CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
 CC invention provides immunogenic compositions and vaccines for dental  
 CC caries. The compositions comprise major histocompatibility complex (MHC)  
 CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
 CC covalently linked with peptide subunits of a glucosyltransferase. The  
 CC compositions are used in a claimed method of eliciting production of an  
 CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be  
 CC prepared synthetically or by recombinant DNA technology. Antibodies  
 CC raised against MHC class II binding fragments of GbpB can be used in  
 CC passive immunisation.  
 XX Sequence 431 AA;  
 SQ  
 Query Match 100.0%; Score 108; DB 7; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGGYGHVAVYTVGGGQIQV 20  
 DB 383 DGGYGHVAVYTVGGGQIQV 402  
 |||||  
 RESULT 5  
 ADD93653  
 ID ADD93653 standard; protein; 431 AA.  
 XX  
 AC ADD93653;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Streptococcus mutans glucan binding protein-B.  
 XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
 XX Streptococcus mutans.  
 XX WO2003075845-A2.  
 XX PD 18-SEP-2003.  
 XX PF 07-MAR-2003; 2003WO-US006962.  
 XX PR 07-MAR-2002; 2002US-0363209P.  
 XX PR 08-AUG-2002; 2002US-0402483P.  
 XX (FORS-) FORSYTH INST.  
 XX Smith DJ, Taubman MA;  
 XX WPI; 2003-845091/78.  
 XX GENBANK; AY046414.  
 XX Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.  
 XX Claim 5; Page 8-9; 49pp; English.  
 XX  
 CC The present sequence is the protein sequence of the glucan binding  
 CC protein-B (GbpB) of Streptococcus mutans strain 5SM3. The sequence  
 CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
 CC invention provides immunogenic compositions and vaccines for dental  
 CC caries. The compositions comprise major histocompatibility complex (MHC)  
 CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
 CC covalently linked with peptide subunits of a glucosyltransferase. The  
 CC compositions are used in a claimed method of eliciting production of an  
 CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be  
 CC prepared synthetically or by recombinant DNA technology. Antibodies  
 CC raised against MHC class II binding fragments of GbpB can be used in  
 CC passive immunisation.  
 XX Sequence 431 AA;  
 SQ  
 Query Match 100.0%; Score 108; DB 7; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGGYGHVAVYTVGGGQIQV 20  
 DB 383 DGGYGHVAVYTVGGGQIQV 402  
 |||||  
 RESULT 5  
 ADD93653  
 ID ADD93653 standard; protein; 431 AA.  
 XX  
 AC ADD93653;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Streptococcus mutans glucan binding protein-B.  
 XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.  
 XX Streptococcus mutans.  
 XX WO2003075845-A2.  
 XX PD 18-SEP-2003.  
 XX PF 07-MAR-2003; 2003WO-US006962.  
 XX PR 07-MAR-2002; 2002US-0363209P.  
 XX PR 08-AUG-2002; 2002US-0402483P.  
 XX (FORS-) FORSYTH INST.  
 XX Smith DJ, Taubman MA;  
 XX WPI; 2003-845091/78.  
 XX GENBANK; AY046414.  
 XX Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.  
 XX Claim 5; Page 8-9; 49pp; English.  
 XX  
 CC The present sequence is the protein sequence of the glucan binding  
 CC protein-B (GbpB) of Streptococcus mutans strain 5SM3. The sequence  
 CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
 CC invention provides immunogenic compositions and vaccines for dental  
 CC caries. The compositions comprise major histocompatibility complex (MHC)  
 CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
 CC covalently linked with peptide subunits of a glucosyltransferase. The  
 CC compositions are used in a claimed method of eliciting production of an  
 CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be  
 CC prepared synthetically or by recombinant DNA technology. Antibodies  
 CC raised against MHC class II binding fragments of GbpB can be used in  
 CC passive immunisation.  
 XX Sequence 431 AA;  
 SQ  
 Query Match 100.0%; Score 108; DB 7; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DGGYGHVAVYTVGGGQIQV 20  
 DB 383 DGGYGHVAVYTVGGGQIQV 402  
 |||||



PP 09-MAR-2004; 2004US-00797821.  
 XX 13-APR-1998; 98US-0081550P.  
 PR 08-JAN-1999; 99US-0115142P.  
 PR 12-APR-1999; 99US-00290049.  
 PR 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 PR 07-MAR-2003; 2003US-00383930.  
 XX (SMIT/) SMITH D J.  
 PA (TAUB/) TAUBMAN M A.  
 XX Smith DJ, Taubman MA;  
 XX WPI; 2005-151644/16.  
 XX New composition comprising a fragment of a glucan binding protein-B (GbpB) that binds to MHC class II protein, and a biocompatible microparticle, useful for producing an antibody (claimed) for immunizing mammals against dental caries.  
 XX Claim 3; SEQ ID NO 33; 73pp; English.  
 XX The invention relates to a composition comprising a fragment of a glucan binding protein-B (GbpB) and a biocompatible microparticle, where the fragment binds to a major histocompatibility complex (MHC) class II protein. The composition is useful for producing an antibody for immunizing mammals against dental caries. This sequence corresponds to a Streptococcus mutans GbpB protein of the invention.  
 XX Sequence 431 AA;  
 Query Match 100.0%; Score 108; DB 9; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 DGGYGHVAVYTGQGGQIQV 20  
 Db 383 DGGYGHVAVYTGQGGQIQV 402  
 RESULT 9  
 AEB91500  
 ID AEB91500 standard; protein; 431 AA.  
 XX AEB91500;  
 XX 20-OCT-2005 (first entry)  
 XX Microbial pathogen adhesin protein sequence, SEQ ID NO:210.  
 XX algorithm; adhesin; pharmaceutical; vaccine; drug screening;  
 KW bordetella pertussis infection; antibacterial; pneumonia;  
 KW antiinflammatory; respiratory-gen.; gastric ulcer; antitumor;  
 KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.  
 XX Streptococcus mutans.  
 OS Streptococcus mutans.  
 XX WO2005076010-A2.  
 PN 18-AUG-2005.  
 PD 07-FEB-2005; 2005WO-IN000037.  
 XX 06-FEB-2004; 2004IN-DE000173.  
 PR 20-JUL-2004; 2004US-0589227P.  
 XX (COUL ) COUNCIL SCI & IND RES SOUTH AFRICA.  
 PA Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramachandran S;  
 PI WPI; 2005-597835/61.  
 XX

PT Computational method for identifying adhesin and adhesin like molecules, comprises computing sequence-based attributes of protein sequences using neural network software and training an artificial neural network.  
 XX Claim 16; SEQ ID NO 210; 402pp; English.  
 XX The present invention relates to a computational method (M1) for identifying adhesin and adhesin-like proteins, by computing the sequence-based attributes of protein sequences using five attribute modules of a neural network software, training an artificial neural network (ANN) for each of the computed five attributes, and identifying the adhesin and adhesin-like proteins having probability of being an adhesin (Pad) as equal or greater than 0.51. Also claimed is a set of 274 annotated genes encoding adhesin and adhesin-like proteins, having 105 fully defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO: 280-384) sequences; and a fully connected multilayer feed forward ANN (I) based on (M1). (M1) is useful for identifying adhesin and adhesin-like proteins, of therapeutic potential, and identifying and short-listing proteins for further testing in development of new vaccine formulations to eliminate diseases caused by various pathogenic organisms. (M1) is useful for identifying putative adhesins that are important in drug discovery and preventing therapeutics for whooping cough, pneumonia, gastric ulcer and urinary tract infections. (M1) identifies adhesins from distantly related organisms, and from bacteria belonging to a wide phylogenetic spectrum. (M1) is capable of predicting adhesive nature of unique proteins. The present sequence is a microbial pathogen adhesin protein sequence.  
 XX Sequence 431 AA;  
 Query Match 100.0%; Score 108; DB 9; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 DGGYGHVAVYTGQGGQIQV 20  
 Db 383 DGGYGHVAVYTGQGGQIQV 402  
 RESULT 10  
 ADD93651  
 ID ADD93651 standard; protein; 432 AA.  
 XX ADD93651;  
 XX 29-JAN-2004 (first entry)  
 DT Streptococcus mutans glucan binding protein-B.  
 DE Glucan binding protein-B; GbpB; vaccine; antitoxins; immunogen.  
 XX Streptococcus mutans.  
 OS Streptococcus mutans.  
 XX WO2003075845-A2.  
 PN 18-SEP-2003.  
 PD 07-MAR-2003; 2003WO-US006962.  
 XX 07-MAR-2002; 2002US-0363209P.  
 PR 08-AUG-2002; 2002US-0402483P.  
 XX (FORS-) FORSYTH INST.  
 PA Smith DJ, Taubman MA;  
 PI WPI; 2003-845091/78.  
 XX GENBANK; AY046412.  
 DR

XX Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.

XX Claim 5; Page 8; 49pp; English.

XX The present sequence is the protein sequence of the glucan binding  
 CC protein-B (GbpB) of Streptococcus mutans strain 15JP2. The sequence  
 CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
 CC invention provides immunogenic compositions and vaccines for dental  
 CC caries. The compositions comprise major histocompatibility complex (MHC)  
 CC class II protein-binding GbpB peptides, especially HLA-binding peptides,  
 CC covalently linked with peptide subunits of a glucosyltransferase. The  
 CC compositions are used in a claimed method of eliciting production of an  
 CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be  
 CC prepared synthetically or by recombinant DNA technology. Antibodies  
 CC raised against MHC class II binding fragments of GbpB can be used in  
 CC passive immunisation.

XX Sequence 432 AA;

Query Match 100.0%; Score 108; DB 7; Length 432;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGGYGHVAVVTGVQGGQIQV 20

DB 384 DGGYGHVAVVTGVQGGQIQV 403

RESULT 11

AD93652 ID ADD93652 standard; protein; 432 AA.

XX AC ADD93652;

XX 29-JAN-2004 (first entry)

XX Streptococcus mutans glucan binding protein-B.

XX Glucan binding protein-B; GbpB; vaccine; anticaries; epitope; immunogen.

XX Streptococcus mutans.

XX WO2003075845-A2.

XX 18-SEP-2003.

XX 07-MAR-2003; 2003WO-US006962.

XX 07-MAR-2002; 2002US-0363209P.

XX 08-AUG-2002; 2002US-0402483P.

XX (FORS-) FORSYTH INST.

XX Smith DJ, Taubman MA;

XX WPI; 2003-845091/78.

XX GENBANK; AY046413.

XX Composition useful as vaccines for dental caries comprises a fragment of  
 PT a glucan binding protein-B binding to a major histocompatibility complex  
 PT class II protein.

XX Claim 5; Page 8; 49pp; English.

XX The present sequence is the protein sequence of the glucan binding  
 CC protein-B (GbpB) of Streptococcus mutans strain 3SN1. The sequence  
 CC includes a number of human leucocyte antigen (HLA)-binding epitopes. The  
 CC invention provides immunogenic compositions and vaccines for dental  
 CC caries. The compositions comprise major histocompatibility complex (MHC)  
 CC class II protein-binding GbpB peptides, especially HLA-binding peptides,

CC covalently linked with peptide subunits of a glucosyltransferase. The  
 CC compositions are used in a claimed method of eliciting production of an  
 CC antibody in a mammal. Diepitopic or multiepitopic polypeptides can be  
 CC prepared synthetically or by recombinant DNA technology. Antibodies  
 CC raised against MHC class II binding fragments of GbpB can be used in  
 CC passive immunisation.

XX Sequence 432 AA;

Query Match 100.0%; Score 108; DB 7; Length 432;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGGYGHVAVVTGVQGGQIQV 20

DB 384 DGGYGHVAVVTGVQGGQIQV 403

RESULT 12

ADX37274

ID ADX37274 standard; protein; 432 AA.

XX AC ADX37274;

XX 21-APR-2005 (first entry)

XX Streptococcus mutant glucan binding protein B variant #3.

XX immunogenicity; immune stimulation; glucan binding protein-B;  
 KW microparticle; major histocompatibility complex; tooth disease.

XX Streptococcus mutans.

XX US2005031633-A1.

XX 10-FEB-2005.

XX 09-MAR-2004; 2004US-00797821.

XX 13-APR-1998; 98US-0081550P.

XX 08-JAN-1999; 99US-0115142P.

XX 12-APR-1999; 99US-00290049.

XX 07-MAR-2002; 2002US-0363209P.

XX 08-AUG-2002; 2002US-0402483P.

XX 07-MAR-2003; 2003US-00383930.

XX (SMIT/) SMITH D J.

XX (TAUB/) TAUBMAN M A.

XX Smith DJ, Taubman MA;

XX WPI; 2005-151644/16.

XX New composition comprising a fragment of a glucan binding protein-B  
 PT (GbpB) that binds to MHC class II protein, and a biocompatible  
 PT microparticle, useful for producing an antibody (claimed) for immunizing  
 PT mammals against dental caries.

XX Claim 3; SEQ ID NO 31; 73pp; English.

XX The invention relates to a composition comprising a fragment of a glucan  
 CC binding protein-B (GbpB) and a biocompatible microparticle, where the  
 CC fragment binds to a major histocompatibility complex (MHC) class II  
 CC protein. The composition is useful for producing an antibody for  
 CC immunizing mammals against dental caries. This sequence corresponds to a  
 CC Streptococcus mutans GbpB protein of the invention.

XX Sequence 432 AA;

Query Match 100.0%; Score 108; DB 9; Length 432;  
 Best Local Similarity 100.0%; Pred. No. 1.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY      1 DGGYGHVAYVTGQGGOIQV 20
Db      384 DGGYGHVAYVTGQGGOIQV 403

RESULT 13
ADX37275
ID ADX37275 standard; protein; 432 AA.
XX
AC ADX37275;
XX
DT 21-APR-2005 (first entry)
XX
DE Streptococcus mutant glucan binding protein B variant #4.
XX
KW immunogenicity; immune stimulation; glucan binding protein-B;
KW microparticle; major histocompatibility complex; tooth disease.
XX
OS Streptococcus mutans.
XX
PN US2005031633-A1.
XX
PD 10-FEB-2005.
XX
PF 09-MAR-2004; 2004US-00797821.
XX
PR 13-APR-1998; 98US-0081550P.
PR 08-JAN-1999; 99US-0115142P.
PR 12-APR-1999; 99US-00290049.
PR 07-MAR-2002; 2002US-0363209P.
PR 08-AUG-2002; 2002US-0402483P.
PR 07-MAR-2003; 2003US-00383930.
XX
PA (SMIT/) SMITH D J.
PA (TAUB/) TAUBMAN M A.
XX
PI Smith DJ, Taubman MA;
XX
DR WPI; 2005-151644/16.
XX
PT New composition comprising a fragment of a glucan binding protein-B
PT (GbpB) that binds to MHC class II protein, and a biocompatible
PT microparticle, useful for producing an antibody (claimed) for immunizing
PT mammals against dental caries.
XX
PS Claim 3; SEQ ID NO 32; 73pp; English.
XX
CC The invention relates to a composition comprising a fragment of a glucan
CC binding protein-B (GbpB) and a biocompatible microparticle, where the
CC fragment binds to a major histocompatibility complex (MHC) class II
CC protein. The composition is useful for producing an antibody for
CC immunizing mammals against dental caries. This sequence corresponds to a
CC Streptococcus mutans GbpB protein of the invention.
XX
SQ Sequence 432 AA;
Query Match 100.0%; Score 108; DB 9; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 DGGYGHVAYVTGQGGOIQV 20
Db      384 DGGYGHVAYVTGQGGOIQV 403

RESULT 14
AEB91642
ID AEB91642 standard; protein; 211 AA.
XX
AC AEB91642;
XX
DT 20-OCT-2005 (first entry)
XX

```

---

```

DE Microbial pathogen adhesin protein sequence, SEQ ID NO:352.
XX algorithm; adhesin; pharmaceutical; vaccine; drug screening;
XX Bordetella pertussis infection; antibacterial; pneumonia;
KW antiinflammatory; respiratory-gen.; gastric ulcer; antiulcer;
KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.
XX
OS Streptococcus mutans.
XX
PN WO2005076010-A2.
XX
PD 18-AUG-2005.
XX
PF 07-FEB-2005; 2005WO-IN000037.
XX
PR 06-FEB-2004; 2004IN-DE000173.
PR 20-JUL-2004; 2004US-0589227P.
XX
PA (COUL ) COUNCIL SCI & IND RES SOUTH AFRICA.
XX
PI Sachdeva G, Kumar K, Jain P, Brahmachari SK, Ramachandran S;
XX WPI; 2005-597835/61.
XX
PT Computational method for identifying adhesin and adhesin like molecules,
PT comprises computing sequence-based attributes of protein sequences using
PT neural network software and training an artificial neural network.
XX
PS Claim 17; SEQ ID NO 352; 402pp; English.
XX
CC The present invention relates to a computational method (M1) for
CC identifying adhesin and adhesin-like proteins, by computing the sequence-
CC based attributes of protein sequences using five attribute modules of a
CC neural network software, training an artificial neural network (ANN) for
CC each of the computed five attributes, and identifying the adhesin and
CC adhesin-like proteins having probability of being an adhesin (Pad) as
CC equal or greater than 0.51. Also claimed is a set of 274 annotated genes
CC encoding adhesin and adhesin-like proteins, having 274 fully defined 162-
CC 1151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical
CC genes encoding adhesin and adhesin-like proteins, having 105 fully
CC defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated
CC adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base
CC pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-
CC like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO:
CC 280-384) sequences; and a fully connected multilayer feed forward ANN (1)
CC based on (M1). (M1) is useful for identifying adhesin and adhesin-like
CC proteins, of therapeutic potential, and identifying and short-listing
CC proteins for further testing in development of new vaccine formulations
CC to eliminate diseases caused by various pathogenic organisms. (M1) is
CC useful for identifying putative adhesins that are important in drug
CC discovery and preventing therapeutics for whooping cough, pneumonia,
CC gastric ulcer and urinary tract infections. (M1) identifies adhesins from
CC distantly related organisms, and from bacteria belonging to a wide
CC phylogenetic spectrum. (M1) is capable of predicting adhesive nature of
CC unique proteins. The present sequence is a microbial pathogen adhesin
CC protein sequence.
XX
SQ Sequence 211 AA;
Query Match 70.8%; Score 76.5; DB 9; Length 211;
Best Local Similarity 71.4%; Pred. No. 0.0031;
Matches 15; Conservative 3; Mismatches 2; Indels 1; Gaps 1;

QY      1 DGGYGHVAYVTGV-QGGOIQV 20
Db      156 DGGYGHVAYVTAVGEDGKQV 176
|||||
|:::|

RESULT 15
AEB91648
ID AEB91648 standard; protein; 544 AA.
XX
AC AEB91648;

```

XX 20-OCT-2005 (first entry)  
DT Microbial pathogen adhesin protein sequence, SEQ ID NO:358.  
DE  
XX  
XX algorithm; adhesin; pharmaceutical; vaccine; drug screening;  
KW bordetella pertussis infection; antibacterial; pneumonia;  
KW antiinflammatory; respiratory-gen.; gastric ulcer; antiulcer;  
KW gastrointestinal-gen.; urinary tract infection; antimicrobial; uropathic.  
XX  
XX Streptococcus mutans.  
OS  
XX WO2005076010-A2.  
PN  
XX  
XX 18-AUG-2005.  
PD  
XX  
XX 07-FEB-2005; 2005WO-IN000037.  
PF  
XX 06-FEB-2004; 2004IN-DE000173.  
PR  
XX 20-JUL-2004; 2004US-0589227P.  
PR  
XX (COUL ) COUNCIL SCI & IND RES SOUTH AFRICA.  
PA  
XX Sachdeva G, Kumar X, Jain P, Brahmachari SK, Ramachandran S;  
PI WPI; 2005-597835/61.  
XX  
DR Computational method for identifying adhesin and adhesin like molecules,  
XX PT comprises computing sequence-based attributes of protein sequences using  
PT neural network software and training an artificial neural network.  
XX  
PS Claim 17; SEQ ID NO 358; 402pp; English.  
XX  
XX The present invention relates to a computational method (M1) for  
CC identifying adhesin and adhesin-like proteins, by computing the sequence-  
CC based attributes of protein sequences using five attribute modules of a  
CC neural network software, training an artificial neural network (ANN) for  
CC each of the computed five attributes, and identifying the adhesin and  
CC adhesin-like proteins having probability of being an adhesin (Pad) as  
CC equal or greater than 0.51. Also claimed is a set of 274 annotated genes  
CC encoding adhesin and adhesin-like proteins, having 274 fully defined 162-  
CC 1151 base pair (SEQ ID NO: 385-658) sequences; a set of 105 hypothetical  
CC genes encoding adhesin and adhesin-like proteins, having 105 fully  
CC defined 306-15876 base pairs (SEQ ID NO: 659-763); a set of 279 annotated  
CC adhesin and adhesin-like proteins, having 279 fully defined 53-3716 base  
CC pair (SEQ ID NO: 1-279); a set of 105 hypothetical adhesin and adhesin-  
CC like proteins, having 105 fully defined 106-5291 base pair (SEQ ID NO:  
CC 280-384) sequences; and a fully connected multilayer feed forward ANN (I)  
CC based on (M1). (M1) is useful for identifying adhesin and adhesin-like  
CC proteins, of therapeutic potential, and identifying and short-listing  
CC proteins for further testing in development of new vaccine formulations  
CC to eliminate diseases caused by various pathogenic organisms. (M1) is  
CC useful for identifying putative adhesins that are important in drug  
CC discovery and preventing therapeutics for whooping cough, pneumonia,  
CC gastric ulcer and urinary tract infections. (M1) identifies adhesins from  
CC distantly related organisms, and from bacteria belonging to a wide  
CC phylogenetic spectrum. (M1) is capable of predicting adhesive nature of  
CC unique proteins. The present sequence is a microbial pathogen adhesin  
CC protein sequence.  
XX  
XX Sequence 544 AA;

Query Match 65.3%; Score 70.5; DB 9; Length 544;  
Best Local Similarity 71.4%; Pred. No. 0.066;  
Matches 15; Conservative 2; Mismatches 3; Indels 1; Gaps 1;  
OY 1 DGGYGHVAVYTVQV-GGQIQV 20  
Db 490 DGGYGHVAVYTVHESNNRIQV 510  
|||||  
:|

Search completed: February 10, 2006, 22:19:40  
Job time : 18.4758 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:20:08 ; Search time 2.35887 Seconds  
(without alignments)  
815.787 Million cell updates/sec

Title: US-10-797-821-21

Perfect score: 108

Sequence: 1 DGGYGHVAVYVGQGOIQV 20

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_80.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	63	58.3	448	2 C95257	choline binding pr
2	63	58.3	448	2 C98122	choline binding pr
3	62.5	57.9	392	2 G95258	secreted 45 kd pro
4	62.5	57.9	392	2 B98124	general stress pro
5	55.5	51.4	255	2 G90061	hypothetical prote
6	52.5	48.6	267	2 F90028	hypothetical prote
7	51.5	47.7	265	2 B89837	hypothetical prote
8	49	45.4	224	2 AB1377	ABC transporter, p
9	49	45.4	224	2 AD1746	ABC transporter, p
10	47	43.5	147	2 AD3081	hypothetical prote
11	47	43.5	147	2 D98205	hypothetical prote
12	47	43.5	433	2 H97846	hypothetical prote
13	47	43.5	808	2 F64914	dimethylsulfoxide
14	47	43.5	808	2 C85764	probable oxidoredu
15	47	43.5	808	2 F90315	probable oxidoredu
16	47	43.5	811	2 AF0680	probable dimethyl
17	46	42.6	134	2 T22275	hypothetical prote
18	46	42.6	222	2 C70020	conserved hypotet
19	46	42.6	469	2 T35670	hypothetical prote
20	46	42.6	2126	2 F90621	probable polyketid
21	45.5	42.1	166	2 C90029	hypothetical prote
22	45.5	42.1	193	1 RDBP74	dihydrofolate redu
23	45	41.7	413	2 G86468	protein F12K21.22
24	45	41.7	445	2 S65734	mitosis-specific c
25	44.5	41.2	442	2 F71930	probable transport
26	44.5	41.2	442	2 A64592	sodium- and chlori
27	44	40.7	124	2 C87439	conserved hypotet
28	44	40.7	231	2 F89856	conserved hypotet
29	44	40.7	309	2 T21195	hypothetical prote

30	44	40.7	373	2 D81242	DnaJ protein NWA02
31	44	40.7	500	2 T03387	polyamine oxidase
32	44	40.7	831	2 G87620	TonB-dependent rec
33	44	40.7	2183	2 T37218	hypothetical prote
34	44	40.7	2478	2 AH2140	polyketide synthas
35	43.5	40.3	1296	2 E81840	transcription-repa
36	43.5	40.3	1379	2 A81102	transcription-repa
37	43	39.8	115	2 H75487	v-type ATP synthas
38	43	39.8	178	2 T07403	TSI-1 protein - to
39	43	39.8	270	2 AC1956	phosphoserase-re
40	43	39.8	363	2 T25278	hypothetical prote
41	43	39.8	395	2 E90465	hypothetical prote
42	43	39.8	496	2 H70568	probable polyketid
43	43	39.8	537	2 T05355	hypothetical prote
44	43	39.8	542	2 T29707	hypothetical prote
45	43	39.8	639	2 S62567	hypothetical prote

ALIGNMENTS

RESULT 1

C95257  
choline binding protein D [imported] - Streptococcus pneumoniae (strain TIGR4)  
C;Species: Streptococcus pneumoniae  
C;Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C;Accession: C95257  
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapf  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A;Reference number: A95000; MUID:21357209; PMID:11463916  
A;Accession: C95257  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-448 <KUR>  
A;Cross-references: UNIPROT:Q9KG22; UNIPARC:UPI0000051B72; GB:A5005672; PIDN:AAK76252.1;  
C;Genetics:  
A;Experimental source: strain TIGR4  
A;Gene: SP2201

Query Match 58.3%; Score 63; DB 2; Length 448;  
Best Local Similarity 57.9%; Pred. No. 0.082;  
Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

Qy 2 GGYGHVAVYVGQGOIQV 20  
Db 127 GTYGHVAVVSNVNGDQIEI 145

RESULT 2

C98122  
choline binding protein D [imported] - Streptococcus pneumoniae (strain R6)  
C;Species: Streptococcus pneumoniae  
C;Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C;Accession: C98122  
R;Hoskins J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Bargett, S.; DeHoff, B.S.; B  
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M  
y, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;  
A;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A;Reference number: A97872; MUID:21429245; PMID:11544234  
A;Accession: C98122  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-448 <KUR>  
A;Cross-references: UNIPROT:Q8DM24; UNIPARC:UPI00000E3768; GB:A5007317; PIDN:AAI00808.1;  
C;Genetics:  
A;Gene: cbppd

Query Match 58.3%; Score 63; DB 2; Length 448;  
Best Local Similarity 57.9%; Pred. No. 0.082;  
Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 DGGYGHVAVVTGVGGQIQV 20  
DB 127 GTYGHVAWNSVMGDQIEI 145

|||||:|:|:|:  
|||:|:|:|:

RESULT 3  
G95258  
secreted 45 kd protein [imported] - Streptococcus pneumoniae (strain TIGR4)  
C/Species: Streptococcus pneumoniae  
C/Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
C/Accession: G95258  
R/Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid  
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,  
nson, T.; Hickey, E.K.; Holt, I.E.  
Science 293, 498-506, 2001  
A/Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,  
A./Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
A/Reference number: A95000; MUID:21357209; PMID:11463916  
A/Accession: G95258  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-392 <KUR>  
A/Cross-references: UNIPROT:Q9N55; UNIPROT:Q8DMY4; UNIPARC:UI0000051B81; GB:AE005672;  
A/Experimental source: strain TIGR4  
C/Genetics:  
A/Gene: SP2216

Query Match 57.9%; Score 62.5; DB 2; Length 392;  
Best Local Similarity 66.7%; Pred. No. 0.084;  
Matches 14; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 1 DGGYGHVAVVTGVGG-QGIQV 20  
DB 338 DGGYGHVAVTVABSTTRIQV 358

|||||:|:|:|:  
|||:|:|:|:

RESULT 4  
B98124  
general stress protein gsp-781 [imported] - Streptococcus pneumoniae (strain R6)  
C/Species: Streptococcus pneumoniae  
C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
C/Accession: B98124  
R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E  
y, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, F.; McAhren, S.; M  
r, P.; Sun, P.M.; Winkler, M.E.  
J. Bacteriol. 183, 5709-5717, 2001  
A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.  
A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
A/Reference number: A97872; MUID:21429245; PMID:11544234  
A/Accession: B98124  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-392 <KUR>  
A/Cross-references: UNIPROT:Q9N55; UNIPROT:Q8DMY4; UNIPARC:UI0000051B81; GB:AE007317;  
C/Genetics:  
A/Gene: gsp-781

Query Match 57.9%; Score 62.5; DB 2; Length 392;  
Best Local Similarity 66.7%; Pred. No. 0.084;  
Matches 14; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 1 DGGYGHVAVVTGVGG-QGIQV 20  
DB 338 DGGYGHVAVTVABSTTRIQV 358

|||||:|:~|:~|:  
|||:|:~|:~|:

RESULT 5  
G90061  
hypothetical protein SA2353 [imported] - Staphylococcus aureus (strain N315)



hypothetical protein RC1176 [imported] - Rickettsia conorii (strain Malish 7)  
C:Species: Rickettsia conorii  
C>Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 09-Jul-2004  
C:Accession: H97846  
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R  
Science 293, 2093-2098, 2001  
A>Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.  
A:Reference number: A97700; MUID:2142074; PMID:11557893  
A:Accession: H97846  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-433 <KUR>  
A:CROSS-references: UNIPROT:Q92GE7; UNIPARC:UPI00000CC00E; GB:AE006914; PIDN:AAL03714.1;  
C:Genetics:  
A:Gene: RC1176

Query Match 43.5%; Score 47; DB 2; Length 433;  
Best Local Similarity 45.0%; Pred. No. 21;  
Matches 9; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 DGGYGHVAVYTGQGQIQV 20  
||| : : : :  
Db 229 DGGYSSGAYTAGIAGTFAV 248

RESULT 13  
F64914  
dimethylsulfoxide reductase (EC 1.8.-.-) chain A2 precursor, anaerobic - Escherichia coli  
C:Species: Escherichia coli  
C>Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 01-Mar-2002  
C:Accession: F64914  
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co  
.A.; Rose, D.J.; Mau, B.; Shao, Y.  
Science 277, 1453-1462, 1997  
A>Title: The complete genome sequence of Escherichia coli K-12.  
A:Reference number: A64720; MUID:97426617; PMID:9278503  
A:Accession: F64914  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-808 <BLAT>  
A:CROSS-references: UNIPARC:UPI00001680BD; GB:AE000254; GB:U00096; NID:g1787862; PIDN:AP  
A:Experimental source: strain K-12, substrain MG1655  
C:Superfamily: trimethylamine-N-oxide reductase  
C:Keywords: 4Fe-4S; iron-sulfur protein; metalloprotein; molybdenum; oxidoreductase  
F:1-46/Domain: signal sequence #status predicted <SIG>  
F:47-808/Product: dimethylsulfoxide reductase chain A2, anaerobic #status predicted <MAT  
F:60,64,68,100/Binding site: iron-sulfur clusters (Cys) (covalent) #status predicted

Query Match 43.5%; Score 47; DB 2; Length 808;  
Best Local Similarity 57.9%; Pred. No. 42;  
Matches 11; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

QY 4 YG----HVAYVTGVOGGQI 18  
||| : : : :  
Db 154 YGNEAVHVLVYGTGVDGGNI 172

RESULT 14  
C85764  
probable oxidoreductase, major subunit Z2576 [imported] - Escherichia coli (strain O157:  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: C85764  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: C85764  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-808 <STO>  
A:CROSS-references: UNIPROT:Q8X4Q1; UNIPARC:UPI00001657DA; GB:AE005174; NID:g12515563; H

A:Experimental source: strain O157:H7, substrain EDL933  
C:Genetics:  
A:Gene: Z2576  
C:Superfamily: trimethylamine-N-oxide reductase

Query Match 43.5%; Score 47; DB 2; Length 808;  
Best Local Similarity 57.9%; Pred. No. 42;  
Matches 11; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

QY 4 YG----HVAYVTGVOGGQI 18  
||| : : : :  
Db 154 YGNEAVHVLVYGTGVDGGNI 172

RESULT 15  
F90915  
probable oxidoreductase major subunit ECs2294 [imported] - Escherichia coli (strain O157:  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C:Accession: F90915  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Iehi, K.; Yokoyama, K.; Han, C.G.;  
Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genom  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: F90915  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-808 <HAY>  
A:CROSS-references: UNIPROT:Q8X4Q1; UNIPARC:UPI00000D0D9B; GB:BA000007; PIDN:BA835717.1;  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: ECs2294  
C:Superfamily: trimethylamine-N-oxide reductase

Query Match 43.5%; Score 47; DB 2; Length 808;  
Best Local Similarity 57.9%; Pred. No. 42;  
Matches 11; Conservative 0; Mismatches 4; Indels 4; Gaps 1;

QY 4 YG----HVAYVTGVOGGQI 18  
||| : : : :  
Db 154 YGNEAVHVLVYGTGVDGGNI 172

Search completed: February 10, 2006, 22:33:09  
JOB time : 5.48387 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:07:53 ; Search time 13.7298 Seconds  
(without alignments)  
1027.730 Million cell updates/sec

Title: US-10-797-821-21

Perfect score: 108

Sequence: 1 DGGYGHVAVYVGQGQIQV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_05.80.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	100.0	431	2	Q938V0_STRMU
2	108	100.0	431	2	Q938V3_STRMU
3	108	100.0	431	2	Q9AG98_STRMU
4	108	100.0	431	2	Q8DWM3_STRMU
5	108	100.0	432	2	Q938V1_STRMU
6	108	100.0	432	2	Q938V2_STRMU
7	76.5	70.8	211	2	Q8DVU8_STRMU
8	70.5	65.3	544	2	Q840V8_STRMU
9	70.5	65.3	544	2	Q840W6_STRMU
10	70.5	65.3	544	2	Q840X3_STRMU
11	70.5	65.3	544	2	Q8DUR7_STRMU
12	65.5	60.6	129	2	Q5M137_STR1
13	65.5	60.6	474	2	Q5M6K4_STR12
14	65.5	60.6	482	2	Q5MSM6_STR12
15	65.5	60.6	485	2	Q5M212_STR11
16	63	58.3	448	2	Q8DMZ4_STRR6
17	63	58.3	448	2	Q9KGZ2_STRPN
18	62.5	57.9	392	2	Q8DMY4_STRR6
19	62.5	57.9	392	2	Q97N55_STRPN
20	62	57.4	169	2	Q8DXT4_STRAS
21	62	57.4	169	2	Q8E3F4_STRAS
22	61	56.5	461	2	Q56SA7_STRTR
23	59.5	55.1	257	2	Q7CCJ3_STAP
24	59.5	55.1	257	2	Q5HLV2_STAEQ
25	57	52.8	398	2	Q5XELL_STRP6
26	57	52.8	398	2	Q9A1Z8_STRPY
27	57	52.8	398	2	Q7CNQ7_STRP8
28	57	52.8	398	2	Q8P318_STRP3
29	56	51.9	328	2	Q6GHC2_STAAR
30	56	51.9	348	2	Q5HFP1_STAAC
31	55.5	51.4	226	2	Q93RG6_STRIT

32	55.5	51.4	255	2	Q53587_STAAR	Q53587 staphylococ
33	55.5	51.4	255	2	Q6G6A8_STAAS	Q6G6A8 staphylococ
34	55.5	51.4	255	2	Q6GDN3_STAAR	Q6GDN3 staphylococ
35	55.5	51.4	255	2	Q5HCY4_STAAC	Q5HCY4 staphylococ
36	55.5	51.4	255	2	Q79ZY3_STAAM	Q79ZY3 staphylococ
37	55.5	51.4	255	2	Q7A2K8_STAAM	Q7A2K8 staphylococ
38	55.5	51.4	255	2	Q7A3D7_STAAN	Q7A3D7 staphylococ
39	55.5	51.4	447	2	Q9AKA4_STRAG	Q9AKA4 streptococc
40	55.5	51.4	447	2	Q8E2H1_STRAS	Q8E2H1 streptococc
41	55.5	51.4	447	2	Q8E7X9_STRAS	Q8E7X9 streptococc
42	53.5	49.5	263	2	Q54487_STACA	Q54487 staphylococ
43	52.5	48.6	156	2	Q5HP26_STAEQ	Q5HP26 staphylococ
44	52.5	48.6	215	2	Q4L8G1_STAAR	Q4L8G1 staphylococ
45	52.5	48.6	267	2	Q5HDQ9_STAAC	Q5HDQ9 staphylococ

ALIGNMENTS

RESULT 1  
Q938V0\_STRMU PRELIMINARY; PRT; 431 AA.  
ID Q938V0; AC Q938V0; DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Glucan-binding protein B.  
OS Streptococcus mutans.  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=1309;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=SSM3;  
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;  
RT "Cloning of the gbpB gene from Streptococcus mutans.";  
RL J. Dent. Res. 79:224-224(2000).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=SSM3;  
RX MEDLINE=21481971; PubMed=11598068;  
RX DOI=10.1128/IAI.69.11.6931-6941.2001;  
RA Mattos-graner R.O., Jin S., King W.F., Chen T., Smith D.J.,  
RA Duncan M.J.;  
RT "Cloning of the Streptococcus mutans gene encoding glucan binding  
RT protein B and analysis of genetic diversity and protein production in  
RT clinical isolates";  
RL Infect. Immun. 69:6931-6941(2001).  
DR ENBL; AY046414; AAK94504.1; -; Genomic\_DNA.  
DR InterPro; IPR007921; CHAP.  
DR InterPro; IPR009148; Siba.  
DR Pfam; PF05257; CHAP; 1.  
DR PRINTS; PR01852; SIBAPROTEIN.  
DR PROSITE; PPS0911; CHAP; 1.  
SQ SEQUENCE 431 AA; 44650 MW; 05D38D8DBC4609F CRC64;  
Query Match 100.0%; Score 108; DB 2; Length 431;  
Best Local Similarity 100.0%; Pred. No. 1.5e-07;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGGYGHVAVYVGQGQIQV 20  
|||||  
Db 383 DGGYGHVAVYVGQGQIQV 402

RESULT 2  
Q938V3\_STRMU PRELIMINARY; PRT; 431 AA.  
ID Q938V3; AC Q938V3; DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Glucan-binding protein B.

```
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5J32;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=5J32;
RX MEDLINE=21481977; PubMed=11598068;
RY DOI=10.1128/JAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
protein B and analysis of genetic diversity and protein production in
clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AY046410; AAK94500.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 431 AA; 44620 MW; 464FE3B563FB7E51 CRC64;

Query Match 100.0%; Score 108; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGGYGHVAVYTVGGGQIQV 20
Db 383 DGGYGHVAVYTVGGGQIQV 402

RESULT 3
Q9AG98_STRMU
ID Q9AG98_STRMU PRELIMINARY; PRT; 431 AA.
AC Q9AG98;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Immunodominant glycoprotein IDG-60 (Glucan-binding protein B).
GN Name=saga;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=21481977; PubMed=11598074;
RY DOI=10.1128/JAI.69.11.6987-6998.2001;
RA Chia J.S., Lee Y.Y., Huang P.T., Chen J.Y.;
RT "Identification of stress-responsive genes in Streptococcus mutans by
differential display reverse transcription-PCR.";
RL Infect. Immun. 69:2493-2501(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
RX MEDLINE=21481977; PubMed=11598074;
RY DOI=10.1128/JAI.69.11.6987-6998.2001;
RA Chia J.S., Chang L.Y., Shun C.T., Chang Y.Y., Chen J.Y.;
RT "A 60-kilodalton immunodominant glycoprotein is essential for cell
wall integrity and the maintenance of cell shape in Streptococcus
mutans.";
RL Infect. Immun. 69:6987-6998(2001).
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=GS-5;
```

```
RA Chia J.-S., Chang L.-Y., Lee Y.-Y., Chen J.-Y.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3VF4;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [5]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3VF4;
RX MEDLINE=21481971; PubMed=11598068;
RY DOI=10.1128/JAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
protein B and analysis of genetic diversity and protein production in
clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AF338445; AAK08104.1; -; Genomic_DNA.
DR EMBL; AY046411; AAK94501.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
SQ SEQUENCE 431 AA; 44592 MW; 3EBE21FC5B47232E CRC64;

Query Match 100.0%; Score 108; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGGYGHVAVYTVGGGQIQV 20
Db 383 DGGYGHVAVYTVGGGQIQV 402

RESULT 4
Q8DWM3_STRMU
ID Q8DWM3_STRMU PRELIMINARY; PRT; 431 AA.
AC Q8DWM3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative secreted antigen gbpB/saga; putative peptidoglycan
hydrolase.
GN Name=gbpB; OrderedLocusNames=SMU.22;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR EMBL; AF014855; AAN57811.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; SIBA.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome.
SQ SEQUENCE 431 AA; 44620 MW; 2D1CA685248CCD3E CRC64;

Query Match 100.0%; Score 108; DB 2; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
```

```
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DGGYGHVAVYTVGGGQIQV 20
    |||||
Db 383 DGGYGHVAVYTVGGGQIQV 402

RESULT 5
Q938V1_STRMU
ID Q938V1_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V1
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=3SN1;
RX MEDLINE=21481971; PubMed=11598068;
DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
protein B and analysis of genetic diversity and protein production in
clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AY046413; AAK94503.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
SQ SEQUENCE 432 AA; 44652 MW; 3F88ECB9A1F3BB4F CRC64;

Query Match 100.0%; Score 108; DB 2; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DGGYGHVAVYTVGGGQIQV 20
    |||||
Db 384 DGGYGHVAVYTVGGGQIQV 403

RESULT 6
Q938V2_STRMU
ID Q938V2_STRMU PRELIMINARY; PRT; 432 AA.
AC Q938V2
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Glucan-binding protein B.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=15JP2;
RA Jin S., Duncan M.J., Taubman M.A., Smith D.J.;
RT "Cloning of the gbpB gene from Streptococcus mutans.";
RL J. Dent. Res. 79:224-224(2000).
RN [2]
RP NUCLEOTIDE SEQUENCE.
```

```
RC STRAIN=15JP2;
RX MEDLINE=21481971; PubMed=11598068;
DOI=10.1128/IAI.69.11.6931-6941.2001;
RA Mattos-Graner R.O., Jin S., King W.F., Chen T., Smith D.J.,
Duncan M.J.;
RT "Cloning of the Streptococcus mutans gene encoding glucan binding
protein B and analysis of genetic diversity and protein production in
clinical isolates.";
RL Infect. Immun. 69:6931-6941(2001).
DR EMBL; AY046412; AAK94502.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS0911; CHAP; 1.
SQ SEQUENCE 432 AA; 44648 MW; E769B2504AEE50E9 CRC64;

Query Match 100.0%; Score 108; DB 2; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.5e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DGGYGHVAVYTVGGGQIQV 20
    |||||
Db 384 DGGYGHVAVYTVGGGQIQV 403

RESULT 7
Q8DVU8_STRMU
ID Q8DVU8_STRMU PRELIMINARY; PRT; 211 AA.
AC Q8DVU8
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein SMU.367.
GN OrderedLocusNames=SMU.367;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=UA159 / ATCC 700610 / Serotype C;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
AJdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
pathogen.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR EMBL; AB014884; AAN58125.1; -; Genomic DNA.
DR GO; GO:0016998; P:cell wall catabolism; IEA.
DR InterPro; IPR007921; CHAP.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01476; LysM; 1.
DR SMART; SM00257; LysM; 1.
DR PROSITE; PS0911; CHAP; 1.
DR Complete proteome.
KW Complete proteome.
SQ SEQUENCE 211 AA; 22470 MW; 7291F26B10CD3937 CRC64;

Query Match 70.8%; Score 76.5; DB 2; Length 211;
Best Local Similarity 71.4%; Pred. No. 0.0038;
Matches 15; Conservative 3; Mismatches 2; Indels 1; Gaps 1;
QY 1 DGGYGHVAVYTVGGGQIQV 20
    |||||
Db 156 DGGYGHVAVYTVGGGQIQV 176

RESULT 8
Q840V8_STRMU
ID Q840V8_STRMU PRELIMINARY; PRT; 544 AA.
```

```
AC Q840V8;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=LM7;
RA Shibata Y.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB108686; BAC75703.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Hypothetical protein.
SQ SEQUENCE 544 AA; 60153 MW; 7BDBE537A3285820 CRC64;

Query Match 65.3%; Score 70.5; DB 2; Length 544;
Best Local Similarity 71.4%; Pred. No. 0.074;
Matches 15; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 DGGYGHVAYVTGVQ-GGQIQV 20
   |||||
Db 490 DGGYGHVAYVTHVESNNRIQV 510

RESULT 9
Q840W6_STRMU
ID Q840W6_STRMU PRELIMINARY; PRT; 544 AA.
AC Q840W6;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT6219;
RA Shibata Y.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB108685; BAC75692.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Hypothetical protein.
SQ SEQUENCE 544 AA; 60066 MW; 370F3D65EBE4190D CRC64;

Query Match 65.3%; Score 70.5; DB 2; Length 544;
Best Local Similarity 71.4%; Pred. No. 0.074;
Matches 15; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 DGGYGHVAYVTGVQ-GGQIQV 20
   |||||
Db 490 DGGYGHVAYVTHVESNNRIQV 510

RESULT 10
Q840X3_STRMU
ID Q840X3_STRMU PRELIMINARY; PRT; 544 AA.
AC Q840X3;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=MT6219;
RA Shibata Y.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB108685; BAC75692.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Hypothetical protein.
SQ SEQUENCE 544 AA; 60066 MW; 370F3D65EBE4190D CRC64;

Query Match 65.3%; Score 70.5; DB 2; Length 544;
Best Local Similarity 71.4%; Pred. No. 0.074;
Matches 15; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 DGGYGHVAYVTGVQ-GGQIQV 20
   |||||
Db 490 DGGYGHVAYVTHVESNNRIQV 510

RESULT 11
Q8DUR7_STRMU
ID Q8DUR7_STRMU PRELIMINARY; PRT; 544 AA.
AC Q8DUR7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein SMU.836.
GN OrderedLocusNames=SMU.836;
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=UA159 / ATCC 700610 / Serotype c;
RX MEDLINE=22295063; PubMed=12397186; DOI=10.1073/pnas.172501299;
RA Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,
RA Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lin S.P.,
RA Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,
RA Ferretti J.J.;
RT "Genome sequence of Streptococcus mutans UA159, a cariogenic dental
pathogen."; Acad. Sci. U.S.A. 99:14434-14439(2002).
RL Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).
DR EMBL; AF014925; AAN58552.1; -; Genomic_DNA.
DR InterPro; IPR007921; CHAP.
DR InterPro; IPR009148; Siba.
DR Pfam; PF05257; CHAP; 1.
DR PRINTS; PR01852; SIBAPROTEIN.
DR PROSITE; PS50911; CHAP; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 544 AA; 60286 MW; D03CAB4331135A46 CRC64;

Query Match 65.3%; Score 70.5; DB 2; Length 544;
Best Local Similarity 71.4%; Pred. No. 0.074;
Matches 15; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

QY 1 DGGYGHVAYVTGVQ-GGQIQV 20
   |||||
Db 490 DGGYGHVAYVTHVESNNRIQV 510

RESULT 12
QSM137_STRTI
```



ID QSM137\_STR1 PRELIMINARY; PRT; 129 AA.  
 AC QSM137;  
 DT 01-FEB-2005 (T-EMBLrel. 29, Created)  
 DT 01-FEB-2005 (T-EMBLrel. 29, Last sequence update)  
 DT 01-FEB-2005 (T-EMBLrel. 29, Last annotation update)  
 DE Glucan binding protein.  
 GN OrderedLocuNames=st0443;  
 OS Streptococcus thermophilus (strain CNRZ 1066).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=299768;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RX PubMed=15543133; DOI=10.1038/nbt1034;  
 RA Bolotin A., Quinkis B., Renault P., Sorokin A., Ehrlich S.D.,  
 RA Kulakauskas S., Lapidus A., Goltzman E., Mazur M., Pusch G.D.,  
 RA Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,  
 RA Ngui K., Masuy D., Hancy F., Burtet S., Boutry M., Delcours J.,  
 RA Goffeau A., Hols P.;  
 RT "Complete sequence and comparative genome analysis of the dairy  
 RT bacterium Streptococcus thermophilus.";  
 RL Nat. Biotechnol. 22:1554-1558(2004).  
 DR EMBL; CP000024; AAV62043.1; -; Genomic\_DNA.  
 DR InterPro; IPR007921; CHAP.  
 DR Pfam; PF05257; Siba.  
 DR PRINTS; PR01852; SIBAPROTEIN.  
 DR PROSITE; PS0911; CHAP; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 129 AA; 13136 MW; 7ADBDA7F41B7D4B3 CRC64;  
  
 Query Match 60.6%; Score 65.5; DB 2; Length 129;  
 Best Local Similarity 70.0%; Pred. No. 0.1;  
 Matches 14; Conservative 0; Mismatches 5; Indels 1; Gaps 1;  
  
 Qy 2 GGYGHVAYVTGVQV-GQIQV 20  
 Db 76 GGYGHVAYVTSAGVNSIQV 95  
  
 RESULT 13  
 QSM6K4\_STR2  
 ID QSM6K4\_STR2 PRELIMINARY; PRT; 474 AA.  
 AC QSM6K4;  
 DT 01-FEB-2005 (T-EMBLrel. 29, Created)  
 DT 01-FEB-2005 (T-EMBLrel. 29, Last sequence update)  
 DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)  
 DE Glucan binding protein (PcsB).  
 GN Names=pcsb; OrderedLocuNames=st0002;  
 OS Streptococcus thermophilus (strain ATCC BAA-250 / LMG 18311).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=264199;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RX PubMed=15543133; DOI=10.1038/nbt1034;  
 RA Bolotin A., Quinkis B., Renault P., Sorokin A., Ehrlich S.D.,  
 RA Kulakauskas S., Lapidus A., Goltzman E., Mazur M., Pusch G.D.,  
 RA Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,  
 RA Ngui K., Masuy D., Hancy F., Burtet S., Boutry M., Delcours J.,  
 RA Goffeau A., Hols P.;  
 RT "Complete sequence and comparative genome analysis of the dairy  
 RT bacterium Streptococcus thermophilus.";  
 RL Nat. Biotechnol. 22:1554-1558(2004).  
 DR EMBL; CP000023; AAV59752.1; -; Genomic\_DNA.

DR EMBL; AY730643; AAW82375.1; -; Genomic\_DNA.  
 DR InterPro; IPR007921; CHAP.  
 DR Pfam; PF05257; Siba.  
 DR PRINTS; PR01852; SIBAPROTEIN.  
 DR PROSITE; PS0911; CHAP; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 474 AA; 48142 MW; 1BAC6A9E0A0A200D CRC64;  
  
 Query Match 60.6%; Score 65.5; DB 2; Length 474;  
 Best Local Similarity 70.0%; Pred. No. 0.36;  
 Matches 14; Conservative 0; Mismatches 5; Indels 1; Gaps 1;  
  
 Qy 2 GGYGHVAYVTGVQV-GQIQV 20  
 Db 421 GGYGHVAYVTSAGVNSIQV 440  
  
 RESULT 14  
 QSM5M6\_STR2  
 ID QSM5M6\_STR2 PRELIMINARY; PRT; 482 AA.  
 AC QSM5M6;  
 DT 01-FEB-2005 (T-EMBLrel. 29, Created)  
 DT 01-FEB-2005 (T-EMBLrel. 29, Last sequence update)  
 DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)  
 DE Cell segregation protein (Cse).  
 GN Names=cse; OrderedLocuNames=st0442;  
 OS Streptococcus thermophilus (strain ATCC BAA-250 / LMG 18311).  
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=264199;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
 RX PubMed=15543133; DOI=10.1038/nbt1034;  
 RA Bolotin A., Quinkis B., Renault P., Sorokin A., Ehrlich S.D.,  
 RA Kulakauskas S., Lapidus A., Goltzman E., Mazur M., Pusch G.D.,  
 RA Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,  
 RA Ngui K., Masuy D., Hancy F., Burtet S., Boutry M., Delcours J.,  
 RA Goffeau A., Hols P.;  
 RT "Complete sequence and comparative genome analysis of the dairy  
 RT bacterium Streptococcus thermophilus.";  
 RL Nat. Biotechnol. 22:1554-1558(2004).  
 DR EMBL; CP000023; AAV60155.1; -; Genomic\_DNA.  
 DR EMBL; AY730642; AAW82373.1; -; Genomic\_DNA.  
 DR GO; GO:0016998; P:cell wall catabolism; IEA.  
 DR InterPro; IPR007921; CHAP.  
 DR InterPro; IPR002482; LysM\_PGBD.  
 DR Pfam; PF05257; CHAP; 1.  
 DR PRINTS; PR01852; SIBAPROTEIN.  
 DR SMART; SM00257; LysM; 1.  
 DR PROSITE; PS0911; CHAP; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 482 AA; 48907 MW; C4C3846BAB7CED7C CRC64;  
  
 Query Match 60.6%; Score 65.5; DB 2; Length 482;  
 Best Local Similarity 70.0%; Pred. No. 0.37;  
 Matches 14; Conservative 0; Mismatches 5; Indels 1; Gaps 1;  
  
 Qy 2 GGYGHVAYVTGVQV-GQIQV 20  
 Db 429 GGYGHVAYVTSAGVNSIQV 448

RESULT 15  
Q5M212\_STRT1  
ID Q5M212\_STRT1 PRELIMINARY; PRT; 485 AA.  
AC Q5M212;  
DT 01-FEB-2005 (TReMBLrel. 29, Created)  
DT 01-FEB-2005 (TReMBLrel. 29, Last sequence update)  
DT 01-FEB-2005 (TReMBLrel. 29, Last annotation update)  
DE Glucan binding protein.  
DE Name=pc9B; OrderedLocusNames=str0022;  
OS Streptococcus thermophilus (strain CNRZ 1066).  
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;  
OC Streptococcus.  
OX NCBI\_TaxID=299768;  
RN [1]  
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].  
RX PubMed=15543133; DOI=10.1038/nbt1034; Sorokin A., Ehrlich S.D.,  
RA Bolotin A., Quinquis B., Renault P., Sorokin A., Ehrlich S.D.,  
RA Kulakauskas S., Lapidus A., Goltzman E., Mazur M., Pusch G.D.,  
RA Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D.,  
RA Ngui K., Masuy D., Hancy F., Burteau S., Boutry M., Delcour J.,  
RA Goffeau A., Hols P.;  
RT "Complete sequence and comparative genome analysis of the dairy  
RT bacterium Streptococcus thermophilus.";  
RL Nat. Biotechnol. 22:1554-1558(2004).  
DR EMBL; CP000024; AAV61641.1; -; Genomic\_DNA.  
DR InterPro; IPR007921; CHAP.  
DR InterPro; IPR009148; Siba.  
DR Pfam; PF05257; CHAP; 1.  
DR PRINTS; PR01852; SIBAPROTEIN.  
DR PROSITE; PS50911; CHAP; 1.  
KW Complete proteome.  
SQ SEQUENCE 485 AA; 49165 MW; 8665E4B1FC6CF7CA CRC64;

Query Match 60.6%; Score 65.5; DB 2; Length 485;  
Best Local Similarity 70.0%; Pred. No. 0.37;  
Matches 14; Conservative 0; Mismatches 5; Indels 1; Gaps 1;  
QY 2 GGYGHVAYVTGVG-GQIQV 20  
DB 432 GGYGHVAYVTSAGNSIQV 451

Search completed: February 10, 2006, 22:31:07  
Job time : 17.7298 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 22:31:39 ; Search time 3.60887 Seconds  
(without alignments)  
458.180 Million cell updates/sec

Title: US-10-797-821-21

Perfect score: 108

Sequence: 1 DGGYGHVAVTVGGGQIQV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : Issued Patents AA:\*
- 1: /cgn2\_6/prodata/1/iaa/5\_COMB.pep.\*
  - 2: /cgn2\_6/prodata/1/iaa/6\_COMB.pep.\*
  - 3: /cgn2\_6/prodata/1/iaa/H\_COMB.pep.\*
  - 4: /cgn2\_6/prodata/1/iaa/PTUS\_COMB.pep.\*
  - 5: /cgn2\_6/prodata/1/iaa/RE\_COMB.pep.\*
  - 6: /cgn2\_6/prodata/1/iaa/backfilese1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	62.5	57.9	392	2	US-09-583-110-4374
2	62.5	57.9	399	2	US-09-107-433-3230
3	59.5	55.1	257	2	US-09-710-279-3244
4	59.5	55.1	264	2	US-09-134-001C-5035
5	59.5	55.1	267	2	US-09-134-001C-4539
6	50.5	46.8	270	2	US-09-134-001C-5441
7	49.5	45.8	161	2	US-09-134-001C-4805
8	49.5	45.8	497	2	US-09-134-000C-5990
9	46	42.6	935	2	US-09-134-000C-6493
10	46	42.6	939	2	US-09-134-000C-5584
11	46	42.6	5087	2	US-09-144-085-1
12	46	42.6	6095	2	US-09-144-085-1
13	45.5	42.1	149	2	US-09-710-279-1682
14	45.5	42.1	157	2	US-09-710-279-2870
15	45	41.7	345	2	US-09-489-039A-12233
16	45	41.7	368	2	US-09-134-000C-4983
17	45	41.7	1281	2	US-09-489-039A-10396
18	44	40.7	232	2	US-09-134-001C-4874
19	44	40.7	414	2	US-10-104-047-2031
20	44	40.7	488	2	US-09-489-039A-13363
21	44	40.7	717	2	US-08-924-629C-5
22	44	40.7	2517	2	US-09-502-540-15380
23	43.5	40.3	677	2	US-09-328-352-7826
24	43	39.8	62	2	US-09-134-000C-3955
25	43	39.8	452	2	US-09-252-991A-19865
26	43	39.8	495	2	US-08-311-731A-3
27	43	39.8	556	2	US-09-348-796A-20229

28	43	39.8	568	2	US-09-328-352-5502	Sequence 5502, Ap
29	43	39.8	605	2	US-09-178-252-4	Sequence 4, Appli
30	43	39.8	605	2	US-09-826-660-4	Sequence 4, Appli
31	43	39.8	655	2	US-09-252-991A-21806	Sequence 21806, A
32	43	39.8	718	1	US-08-731-079A-1	Sequence 1, Appli
33	43	39.8	1048	2	US-09-921-099A-11	Sequence 11, Appli
34	43	39.8	1148	1	US-08-349-867-23	Sequence 23, Appli
35	43	39.8	1148	1	US-08-349-867-27	Sequence 27, Appli
36	43	39.8	1148	1	US-08-239-476-23	Sequence 23, Appli
37	43	39.8	1148	1	US-08-239-476-27	Sequence 27, Appli
38	43	39.8	1148	1	US-08-598-305A-23	Sequence 23, Appli
39	43	39.8	1148	1	US-08-598-305A-27	Sequence 27, Appli
40	43	39.8	1148	1	US-08-598-305A-35	Sequence 35, Appli
41	43	39.8	1148	1	US-08-598-305A-37	Sequence 37, Appli
42	43	39.8	1148	1	US-08-598-305A-38	Sequence 38, Appli
43	43	39.8	1148	1	US-08-639-923A-23	Sequence 23, Appli
44	43	39.8	1148	1	US-08-639-923A-27	Sequence 27, Appli
45	43	39.8	1148	1	US-08-639-923A-35	Sequence 35, Appli

ALIGNMENTS

RESULT 1  
US-09-583-110-4374  
; Sequence 4374, Application US/09583110  
; Patent No. 6699703  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al.  
; TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus  
; FILE REFERENCE: PATH00-07A  
; CURRENT APPLICATION NUMBER: US/09/583,110  
; CURRENT FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/107,433  
; PRIOR FILING DATE: 1998-06-30  
; PRIOR APPLICATION NUMBER: US 60/085,131  
; PRIOR FILING DATE: 1998-05-12  
; PRIOR APPLICATION NUMBER: US 60/051,553  
; PRIOR FILING DATE: 1997-07-02  
; NUMBER OF SEQ ID NOS: 5322  
; SEQ ID NO 4374  
; LENGTH: 392  
; TYPE: PRT  
; ORGANISM: Streptococcus pneumoniae  
US-09-583-110-4374

Query Match 57.9%; Score 62.5; DB 2; Length 392;  
Best Local Similarity 66.7%; Pred. No. 0.22;  
Matches 14; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 1 DGGYGHVAVTVGGQ-GQIQV 20  
||||||| : : :  
Db 338 DGGYGHVAVTVGGQ-GQIQV 358

RESULT 2  
US-09-107-433-3230  
; Sequence 3230, Application US/09107433  
; Patent No. 6800744  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID  
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE  
; PNEUMONIAE  
; NUMBER OF SEQUENCES: 5206  
; CORRESPONDENCE ADDRESS:  
; ADDRESSED: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354

```

,
, COMPUTER READABLE FORM:
,
, MEDIUM TYPE: CD/ROM ISO9660
,
, COMPUTER: <Unknown>
,
, OPERATING SYSTEM: <Unknown>
,
, SOFTWARE: <Unknown>
,
, CURRENT APPLICATION DATA:
,
, APPLICATION NUMBER: US/09/107,433
,
, FILING DATE: 30-Jun-1998
,
, PRIOR APPLICATION DATA:
,
, APPLICATION NUMBER: 60/ 085131
,
, FILING DATE: May 12, 1998
,
, APPLICATION NUMBER: 60/051553
,
, FILING DATE: July 2, 1997
,
, ATTORNEY/AGENT INFORMATION:
,
, NAME: Ariniello, Pamela Deneke
,
, REGISTRATION NUMBER: 40, 489
,
, REFERENCE/DOCKET NUMBER: GTC-011
,
, TELECOMMUNICATION INFORMATION:
,
, TELEPHONE: (781)893-5007
,
, TELEFAX: (781)893-8277
,
, INFORMATION FOR SEQ ID NO: 3230:
,
, SEQUENCE CHARACTERISTICS:
,
, LENGTH: 399 amino acids
,
, TYPE: amino acid
,
, TOPOLOGY: linear
,
, MOLECULE TYPE: protein
,
, HYPOTHETICAL: YES
,
, ORIGINAL SOURCE:
,
, ORGANISM: Streptococcus pneumoniae
,
, FEATURE:
,
, NAME/KEY: misc feature
,
, LOCATION: (B) LOCATION 1...399
,
, SEQUENCE DESCRIPTION: SEQ ID NO: 3230:
,
, US-09-107-433-3230
,

```

```

Query Match      57.9%; Score 62.5; DB 2; Length 399;
Best Local Similarity 66.7%; Pred.No. 0.22;
Matches 14; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY 1 DGGYGHVAVYTGQG-QGIQV 20
    ||||| ||| | : |||
Db 345 DGGYGHVAVYAVESTTRIOV 365

```

```

RESULT 3
US-09-710-279-3244
; Sequence 3244, Application US/09710279
; Patent No. 6703492
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/09/710,279
; CURRENT FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3244
; LENGTH: 257
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-09-710-279-3244

```

```

Query Match          55.1%; Score 59.5; DB 2; Length 257;
Best Local Similarity 65.0%; Pred. No. 0.38;
Matches 13; Conservative 1; Mismatches 5; Indels 1; Gaps 1;
Qy      2 GSYGHVAYVTGV-QGGQIQV 20
      ||||| ||||| |||||
      :

```

Db 209 GAYGHVAYVEGVNSNGSIRV 228

RESULT 4  
 US-09-134-001C-5035  
 ; Sequence 5035, Application US/09134001C  
 ; Patent No. 6380370  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lynn Doucette-Stamm et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 ; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: GFC-007  
 ; CURRENT APPLICATION NUMBER: US/09/134,001C  
 ; CURRENT FILING DATE: 1998-08-13  
 ; PRIOR APPLICATION NUMBER: US 60/064,964  
 ; PRIOR FILING DATE: 1997-11-08  
 ; PRIOR APPLICATION NUMBER: US 60/055,779  
 ; PRIOR FILING DATE: 1997-08-14  
 ; NUMBER OF SEQ ID NOS: 5674  
 ; SEQ ID NO 5035  
 ; LENGTH: 264  
 ; TYPE: PRT  
 ; ORGANISM: Staphylococcus epidermidis  
 US-09-134-001C-5035

Query Match	55.1%;	Score 59.5;	DB 2;	Length 264;
Best Local Similarity	65.0%;	Pred. No. 0.39;		
Matches 13;	Conservative	1;	Mismatches 5;	Indels 1;
Gaps 1;				

db 216 GAYGHVAYVEGVNSNGSIRV 235

```

RESULT 5
US-09-134-001C-4539
; Sequence 4539, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GFC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4539
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4539

```

Query Match	55.1%	Score 59.5;	DB 2;	Length 267;
Best Local Similarity	65.0%;			
Matches 13: Conservative		Pred. No. 0.39;		
		1: Mismatches	5: Indels	Gaps 1:

219 GAYGHVAYVEGVNSNGSIRV 238

RESULT 6  
US-09-134-001C-5441  
; Sequence 5441, Application US/09134001C  
; Patent No. 6380370  
; GENERAL INFORMATION:  
; APPLICANT: Lynn Doucette-Stamm et al  
; TITLE OF INVENTION: NUCLEIC ACID AND A  
; TITLE OF INVENTION: EPIDERMIS FOR D

```

; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5441
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5441

Query Match      46.8%; Score 50.5; DB 2; Length 270;
Best Local Similarity 55.0%; Pred. No. 8.6;
Matches 11; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

QY      2 GGYGHVAYVTGVG-QGIQV 20
Db      222 GPGYHVAYVERINGDGSILI 241

RESULT 7
US-09-134-001C-4805
; Sequence 4805, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4805
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4805

Query Match      45.8%; Score 49.5; DB 2; Length 161;
Best Local Similarity 50.0%; Pred. No. 6.8;
Matches 10; Conservative 4; Mismatches 5; Indels 1; Gaps 1;

QY      2 GGYGHVAYVTGV-QGGQIQV 20
Db      108 GKGYHVAYIETINEDGSMQV 127

RESULT 8
US-09-134-000C-5990
; Sequence 5990, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5990
; LENGTH: 497
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-144-085-1
```

```

US-09-134-000C-5990

Query Match      45.8%; Score 49.5; DB 2; Length 497;
Best Local Similarity 52.6%; Pred. No. 24;
Matches 10; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

QY      2 GGYGHVAYVTGVGGQIQV 20
Db      445 GGV-HTVLVTGVSQSVQI 462

RESULT 9
US-09-134-000C-6493
; Sequence 6493, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6493
; LENGTH: 935
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-6493

Query Match      42.6%; Score 46; DB 2; Length 935;
Best Local Similarity 42.1%; Pred. No. 1.6e+02;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY      1 DGGYGHVAYVTGVGGQIQ 19
Db      885 DPTYGHTGVYGLNNGRIQ 903

RESULT 10
US-09-134-000C-5584
; Sequence 5584, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5584
; LENGTH: 939
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-134-000C-5584

Query Match      42.6%; Score 46; DB 2; Length 939;
Best Local Similarity 42.1%; Pred. No. 1.6e+02;
Matches 8; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY      1 DGGYGHVAYVTGVGGQIQ 19
Db      889 DPTYGHTGVYGLNNGRIQ 907

RESULT 11
US-09-144-085-1
```

; Sequence 1, Application US/09144085  
; Patent No. 6280999  
; GENERAL INFORMATION:  
; APPLICANT: Gustafsson, Claes  
; APPLICANT: Betlach, Mary C.  
; APPLICANT: Ashley, Gary  
; APPLICANT: Julien, Bryan  
; APPLICANT: Ziermann, Rainer  
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA  
; FILE REFERENCE: 30062-20020.20  
; CURRENT APPLICATION NUMBER: US/09/144,085  
; CURRENT FILING DATE: 1998-08-31  
; EARLIER APPLICATION NUMBER: 09/010,809  
; EARLIER FILING DATE: 1998-01-22  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; TYPE: PRT  
; LENGTH: 5087  
; ORGANISM: Sorangium cellulosum  
US-09-144-085-1

Query Match 42.6%; Score 46; DB 2; Length 5087;  
Best Local Similarity 62.5%; Pred. No. 1.1e+03;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 5 GHVAYVTGQGGOIQV 20  
| | | | | | | | | |  
Db 67 GRVAYVLGLQGAITV 82

RESULT 12  
US-09-144-085-2  
; Sequence 2, Application US/09144085  
; Patent No. 6280999  
; GENERAL INFORMATION:  
; APPLICANT: Gustafsson, Claes  
; APPLICANT: Betlach, Mary C.  
; APPLICANT: Ashley, Gary  
; APPLICANT: Julien, Bryan  
; APPLICANT: Ziermann, Rainer  
; TITLE OF INVENTION: SORANGIUM POLYKETIDE SYNTHASES AND ENCODING DNA  
; FILE REFERENCE: 30062-20020.20  
; CURRENT APPLICATION NUMBER: US/09/144,085  
; CURRENT FILING DATE: 1998-08-31  
; EARLIER APPLICATION NUMBER: 09/010,809  
; EARLIER FILING DATE: 1998-01-22  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 6095  
; TYPE: PRT  
; ORGANISM: Sorangium cellulosum  
US-09-144-085-2

Query Match 42.6%; Score 46; DB 2; Length 6095;  
Best Local Similarity 62.5%; Pred. No. 1.4e+03;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 5 GHVAYVTGQGGOIQV 20  
| | | | | | | | | |  
Db 185 GRVAYVLGLQGAITV 200

RESULT 13  
US-09-710-279-1682  
; Sequence 1682, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS

; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1682  
; LENGTH: 149  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-09-710-279-1682

Query Match 42.1%; Score 45.5; DB 2; Length 149;  
Best Local Similarity 47.6%; Pred. No. 24;  
Matches 10; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 1 DGGYGHVAVVTGVQG-GQIQV 20  
| | | | | | | | | |  
Db 106 EGAFGHVAFVESVNNDGSITV 126

RESULT 14  
US-09-710-279-2870  
; Sequence 2870, Application US/09710279  
; Patent No. 6703492  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/09/710,279  
; CURRENT FILING DATE: 2000-11-09  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2870  
; LENGTH: 157  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-09-710-279-2870

Query Match 42.1%; Score 45.5; DB 2; Length 157;  
Best Local Similarity 47.6%; Pred. No. 26;  
Matches 10; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

QY 1 DGGYGHVAVVTGVQG-GQIQV 20  
| | | | | | | | | |  
Db 106 EGAFGHVAFVESVNNDGSITV 126

RESULT 15  
US-09-489-039A-12233  
; Sequence 12233, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 12233  
; LENGTH: 345

; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-12233

Query Match 41.7%; Score 45; DB 2; Length 345;  
Best Local Similarity 44.4%; Pred. No. 75;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

Qy 2 GGYGHVAYVTGVQGGQIQ 19  
||| |::|  
Db 205 GGNATAAYTTGINAGKMQ 222

Search completed: February 10, 2006, 22:36:12  
Job time : 4.60887 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:14:04 ; Search time 11.7339 Seconds  
(without alignments)  
712.176 Million cell updates/sec

Title: US-10-797-821-21

Perfect score: 108

Sequence: 1 DGGYGHVAVYTVGGGQIQV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications AA.Main.\*
- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	100.0	20	4	US-10-383-930-21
2	108	100.0	20	5	US-10-797-821-21
3	108	100.0	431	4	US-10-383-930-29
4	108	100.0	431	4	US-10-383-930-30
5	108	100.0	431	4	US-10-383-930-33
6	108	100.0	431	5	US-10-797-821-29
7	108	100.0	431	5	US-10-797-821-30
8	108	100.0	431	5	US-10-797-821-33
9	108	100.0	432	4	US-10-383-930-31
10	108	100.0	432	4	US-10-383-930-32
11	108	100.0	432	5	US-10-797-821-31
12	108	100.0	432	5	US-10-797-821-32
13	63	58.3	448	5	US-10-472-928-4622
14	62.5	57.9	392	5	US-10-472-928-4652
15	62.5	57.9	399	5	US-10-617-320-3230
16	59.5	55.1	257	5	US-10-470-048B-362
17	59.5	55.1	264	4	US-10-724-972A-6539
18	59.5	55.1	267	4	US-10-724-972A-5110
19	57	52.8	398	5	US-10-474-792-600
20	55.5	51.4	261	5	US-10-470-048B-65
21	52.5	48.6	267	5	US-10-470-048B-74
22	51.5	47.7	265	5	US-10-470-048B-89
23	50.5	46.8	266	5	US-10-470-048B-285
24	50.5	46.8	270	4	US-10-724-972A-7095
25	50	46.3	226	4	US-10-374-780A-904
26	50	46.3	226	4	US-10-412-699B-1213
27	50	46.3	226	4	US-10-437-963-119616

28	49.5	45.8	161	4	US-10-724-972A-5591	Sequence 5591, Ap
29	49	45.4	224	4	US-10-282-122A-60391	Sequence 60391, A
30	49	45.4	424	4	US-10-156-761-8087	Sequence 8087, Ap
31	48	44.4	99	4	US-10-767-701-61612	Sequence 61612, A
32	48	44.4	108	4	US-10-425-115-275803	Sequence 275803, A
33	48	44.4	135	4	US-10-437-963-129038	Sequence 129038, A
34	48	44.4	336	4	US-10-425-115-286863	Sequence 286863, A
35	48	44.4	374	5	US-10-474-792-598	Sequence 598, App
36	47	43.5	69	4	US-10-781-014-648	Sequence 203635, App
37	47	43.5	203	4	US-10-437-963-203635	Sequence 203635, App
38	47	43.5	264	3	US-09-738-626-3995	Sequence 3995, Ap
39	47	43.5	630	5	US-10-450-763-35420	Sequence 35420, A
40	47	43.5	835	5	US-10-450-763-59154	Sequence 59154, A
41	46	42.6	90	4	US-10-424-599-213697	Sequence 213697, A
42	46	42.6	100	4	US-10-767-701-42596	Sequence 42596, A
43	46	42.6	144	4	US-10-425-114-41744	Sequence 41744, A
44	46	42.6	255	4	US-10-425-115-252874	Sequence 252874, A
45	46	42.6	679	4	US-10-437-963-151293	Sequence 151293, A

ALIGNMENTS

RESULT 1  
US-10-383-930-21  
; Sequence 21, Application US/10383930  
; Publication No. US20040127400A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J  
; APPLICANT: Taubman, Martin A  
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein  
; FILE REFERENCE: 25669-018  
; CURRENT APPLICATION NUMBER: US/10/383,930  
; CURRENT FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 21  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans  
; US-10-383-930-21

Query Match 100.0%; Score 108; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred No. 5.3e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGGYGHVAVYTVGGGQIQV 20  
| | | | | | | | | | | | | | | | | | | | | |  
Db 1 DGGYGHVAVYTVGGGQIQV 20

RESULT 2  
US-10-797-821-21  
; Sequence 21, Application US/10797821  
; Publication No. US20050031633A1  
; GENERAL INFORMATION:  
; APPLICANT: Smith, Daniel J.  
; APPLICANT: Taubman, Martin A.  
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens  
; FILE REFERENCE: 25669-020  
; CURRENT APPLICATION NUMBER: US/10/797,821  
; CURRENT FILING DATE: 2004-03-09  
; PRIOR APPLICATION NUMBER: 10/383,930  
; PRIOR FILING DATE: 2003-03-07  
; PRIOR APPLICATION NUMBER: 60/363,209  
; PRIOR FILING DATE: 2002-03-07  
; PRIOR APPLICATION NUMBER: 60/402,483  
; PRIOR FILING DATE: 2002-08-08  
; PRIOR APPLICATION NUMBER: 09/290,049

```
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 21
; LENGTH: 20
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: GbpB peptide
US-10-797-821-21
```

```
Query Match      100.0%; Score 108; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 5.3e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DGGYGHVAVYTVGGGQIQV 20
Db 1 DGGYGHVAVYTVGGGQIQV 20
|||||
```

## RESULT 3

```
US-10-383-930-29
; Sequence 29, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-29
```

```
Query Match      100.0%; Score 108; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DGGYGHVAVYTVGGGQIQV 20
Db 383 DGGYGHVAVYTVGGGQIQV 402
|||||
```

## RESULT 4

```
US-10-383-930-30
; Sequence 30, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
```

```
; SEQ ID NO 30
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-30
```

```
Query Match      100.0%; Score 108; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DGGYGHVAVYTVGGGQIQV 20
Db 383 DGGYGHVAVYTVGGGQIQV 402
|||||
```

## RESULT 5

```
US-10-383-930-33
; Sequence 33, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J
; APPLICANT: Taubman, Martin A
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 33
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-33
```

```
Query Match      100.0%; Score 108; DB 4; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 DGGYGHVAVYTVGGGQIQV 20
Db 383 DGGYGHVAVYTVGGGQIQV 402
|||||
```

## RESULT 6

```
US-10-797-821-29
; Sequence 29, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 29
```

```
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-29

Query Match      100.0%; Score 108; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGGYGHVAVYTVGGGQIQV 20
    |||||
Db 383 DGGYGHVAVYTVGGGQIQV 402

RESULT 7
US-10-797-821-30
; Sequence 30, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-30

Query Match      100.0%; Score 108; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGGYGHVAVYTVGGGQIQV 20
    |||||
Db 383 DGGYGHVAVYTVGGGQIQV 402

RESULT 8
US-10-797-821-33
; Sequence 33, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 30
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-33

Query Match      100.0%; Score 108; DB 5; Length 431;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGGYGHVAVYTVGGGQIQV 20
    |||||
Db 383 DGGYGHVAVYTVGGGQIQV 402

RESULT 9
US-10-383-930-31
; Sequence 31, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-31

Query Match      100.0%; Score 108; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGGYGHVAVYTVGGGQIQV 20
    |||||
Db 384 DGGYGHVAVYTVGGGQIQV 403

RESULT 10
US-10-383-930-32
; Sequence 32, Application US/10383930
; Publication No. US20040127400A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Immunogenicity of Glucan Binding Protein
; FILE REFERENCE: 25669-018
; CURRENT APPLICATION NUMBER: US/10/383,930
; CURRENT FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-383-930-32
```

```
; ORGANISM: Streptococcus mutans
US-10-383-930-32

Query Match      100.0%; Score 108; DB 4; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGGYGHVAVVTGQGQIQV 20
   |||||
Db 384 DGGYGHVAVVTGQGQIQV 403

RESULT 11
US-10-797-821-31
; Sequence 31, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 31
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-31

Query Match      100.0%; Score 108; DB 5; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGGYGHVAVVTGQGQIQV 20
   |||||
Db 384 DGGYGHVAVVTGQGQIQV 403

RESULT 12
US-10-797-821-32
; Sequence 32, Application US/10797821
; Publication No. US20050031633A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Daniel J.
; APPLICANT: Taubman, Martin A.
; TITLE OF INVENTION: Glucan Binding Protein and Glycosyltransferase Immunogens
; FILE REFERENCE: 25669-020
; CURRENT APPLICATION NUMBER: US/10/797,821
; CURRENT FILING DATE: 2004-03-09
; PRIOR APPLICATION NUMBER: 10/383,930
; PRIOR FILING DATE: 2003-03-07
; PRIOR APPLICATION NUMBER: 60/363,209
; PRIOR FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/402,483
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: 09/290,049
; PRIOR FILING DATE: 1999-04-12
; PRIOR APPLICATION NUMBER: 60/081,550
; PRIOR FILING DATE: 1998-04-13
; ORGANISM: Streptococcus mutans

; PRIOR APPLICATION NUMBER: 60/115,142
; PRIOR FILING DATE: 1999-01-08
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 432
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-10-797-821-32

Query Match      100.0%; Score 108; DB 5; Length 432;
Best Local Similarity 100.0%; Pred. No. 1.4e-07;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DGGYGHVAVVTGQGQIQV 20
   |||||
Db 384 DGGYGHVAVVTGQGQIQV 403

RESULT 13
US-10-472-928-4622
; Sequence 4622, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: Seqwin99, version 1.03
; SEQ ID NO 4622
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; OTHER INFORMATION: choline binding protein D (cbpD)
; OTHER INFORMATION: Cellular location: membrane
; OTHER INFORMATION: Feature of note: WY motif, RGD motif
; OTHER INFORMATION: Similar to strain R6 sequence 15904047 (0.E+01)
US-10-472-928-4622

Query Match      58.3%; Score 63; DB 5; Length 448;
Best Local Similarity 57.9%; Pred. No. 0.68;
Matches 11; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 2 GGYGHVAVVTGQGQIQV 20
   |||||
Db 127 GTYGHVAVVNSVWGDQIEI 145

RESULT 14
US-10-472-928-4652
; Sequence 4652, Application US/10472928
; Publication No. US20050020813A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: THE INSTITUTE FOR GENOMIC RESEARCH
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: P026926WO
; CURRENT APPLICATION NUMBER: US/10/472,928
; CURRENT FILING DATE: 2003-09-26
; PRIOR APPLICATION NUMBER: GB-0107658.7
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 4979
; SOFTWARE: Seqwin99, version 1.03
; SEQ ID NO 4652
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
```

```
/
/
/ FEATURE:
/ OTHER INFORMATION: secreted 45 kd protein (usp45)
/ OTHER INFORMATION: Cellular location: outside
/ OTHER INFORMATION: Similar to strain R6 sequence 15904062 (0.E+01)
US-10-472-928-4652
Query Match      57.9%; Score 62.5; DB 5; Length 392;
Best Local Similarity 66.7%; Pred.No. 0.7;
Matches 14; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

QY      1 DGGYGHVAVVTGVQG-GQIQV 20
Db      338 DGGYGHVAVVTAVESTTRIQV 358

RESULT 15
US-10-617-320-3230
; Sequence 3230, Application US/10617320
; Publication No. US20050136404A1
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE FOR DIAGNOSTIC
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/617,320
; FILING DATE: 10-Jul-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3230:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...399
; SEQUENCE DESCRIPTION: SEQ ID NO: 3230:
US-10-617-320-3230

Query Match      57.9%; Score 62.5; DB 5; Length 399;
Best Local Similarity 66.7%; Pred.No. 0.71;
Matches 14; Conservative 2; Mismatches 4; Indels 1; Gaps 1;
```

QY 1 DGGYGHVAVVTGVQG-GQIQV 20  
Db 345 DGGYGHVAVVTAVESTTRIQV 365

Search completed: February 10, 2006, 23:25:08  
Job time : 12.7339 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 10, 2006, 23:15:51 ; Search time 0.866935 Seconds  
(without alignments)  
302.734 Million cell updates/sec

Title: US-10-797-821-21

Perfect score: 108

Sequence: 1 DGGYGHVAVYVGQGGIQV 20

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 97014 seqs, 13122538 residues

Total number of hits satisfying chosen parameters: 97014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Published Applications\_AA\_New.\*
- 1: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep.\*
  - 2: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep.\*
  - 3: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep.\*
  - 4: /cgn2\_6/ptodata/1/pubppaa/PCT\_NEW\_PUB.pep.\*
  - 5: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep.\*
  - 6: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep.\*
  - 7: /cgn2\_6/ptodata/1/pubppaa/US11\_NEW\_PUB.pep.\*
  - 8: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	108	100.0	431	7	US-11-052-554A-210 Sequence 210, App
2	76.5	70.8	211	7	US-11-052-554A-352 Sequence 352, App
3	70.5	65.3	544	7	US-11-052-554A-358 Sequence 358, App
4	63	58.3	448	7	US-11-052-554A-230 Sequence 230, App
5	59.5	55.1	257	6	US-10-793-626-3244 Sequence 3244, App
6	57	52.8	398	7	US-11-052-554A-252 Sequence 252, App
7	50.5	46.8	270	6	US-10-485-517-413 Sequence 413, App
8	45.5	42.1	149	6	US-10-793-626-1682 Sequence 1682, App
9	45.5	42.1	157	6	US-10-793-626-2870 Sequence 2870, App
10	44	40.7	344	6	US-10-467-657-8284 Sequence 8284, App
11	43.5	40.3	1234	6	US-10-467-657-4224 Sequence 4224, App
12	43	39.8	183	7	US-11-055-822-84 Sequence 84, Appl
13	43	39.8	639	7	US-11-052-554A-165 Sequence 165, App
14	42.5	39.4	756	6	US-10-467-657-8694 Sequence 8694, App
15	42	38.9	204	6	US-10-467-657-6872 Sequence 6872, App
16	42	38.9	219	6	US-10-467-657-6864 Sequence 6864, App
17	42	38.9	238	6	US-10-467-657-7864 Sequence 7864, App
18	42	38.9	358	6	US-10-467-657-7030 Sequence 7030, App
19	42	38.9	365	7	US-11-052-554A-223 Sequence 223, App
20	42	38.9	366	6	US-10-467-657-7024 Sequence 7024, App
21	42	38.9	366	6	US-10-467-657-7964 Sequence 7964, App
22	41.5	38.4	143	6	US-10-793-626-112 Sequence 112, App
23	41.5	38.4	143	6	US-10-793-626-1940 Sequence 1940, App
24	41.5	38.4	330	6	US-10-485-517-415 Sequence 415, App
25	41	38.0	267	6	US-10-454-437-284 Sequence 284, App

26	41	38.0	365	6	US-10-467-657-3846 Sequence 3846, Ap
27	41	38.0	374	7	US-11-052-554A-256 Sequence 256, App
28	41	38.0	495	6	US-10-770-726-81 Sequence 81, Appl
29	41	38.0	914	7	US-11-052-554A-160 Sequence 160, App
30	41	38.0	946	7	US-11-207-626A-44 Sequence 44, Appl
31	41	38.0	952	7	US-11-207-626A-45 Sequence 45, Appl
32	41	38.0	958	7	US-11-207-626A-43 Sequence 43, Appl
33	40	37.0	670	7	US-11-150-883-51 Sequence 51, Appl
34	40	37.0	685	7	US-11-150-883-51 Sequence 51, Appl
35	40	37.0	1892	7	US-11-075-185-6 Sequence 5, Appli
36	40	37.0	3655	7	US-11-075-185-5 Sequence 4, Appli
37	40	37.0	3689	7	US-11-075-185-4 Sequence 49, Appl
38	40	37.0	7968	7	US-11-143-980-49 Sequence 209, App
39	39.5	36.6	280	7	US-11-020-602-209 Sequence 374, App
40	39.5	36.6	619	6	US-10-485-517-374 Sequence 208, App
41	39.5	36.6	1052	7	US-11-020-602-208 Sequence 1052, Ap
42	39	36.1	655	6	US-10-793-626-1052 Sequence 1400, Ap
43	39	36.1	655	6	US-10-793-626-1400 Sequence 42, Appl
44	39	36.1	2367	7	US-11-051-453-42 Sequence 7, Appli
45	39	36.1	3507	7	US-11-075-185-7

ALIGNMENTS

RESULT 1  
US-11-052-554A-210  
; Sequence 210, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR FILING DATE: 2004-02-06  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 210  
; LENGTH: 431  
; TYPE: PRT  
; ORGANISM: Streptococcus mutans UA159  
US-11-052-554A-210

Query Match 100.0%; Score 108; DB 7; Length 431;  
Best Local Similarity 100.0%; Pred. No. 9.7e-09;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DGGYGHVAVYVGQGGIQV 20  
Db 383 DGGYGHVAVYVGQGGIQV 402

RESULT 2  
US-11-052-554A-352  
; Sequence 352, Application US/11052554A  
; Publication No. US20050288866A1  
; GENERAL INFORMATION:  
; APPLICANT: Sachdeva, et al.  
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE  
; FILE REFERENCE: 30853/40359A  
; CURRENT APPLICATION NUMBER: US/11/052,554A  
; CURRENT FILING DATE: 2005-02-07  
; PRIOR FILING DATE: 2005-02-07  
; PRIOR FILING DATE: 2004-07-20  
; PRIOR FILING DATE: 2004-02-06  
; NUMBER OF SEQ ID NOS: 763





; Sequence 413, Application US/10485517  
; Publication No. US20050256299A1  
; GENERAL INFORMATION:  
; APPLICANT: University of Sheffield  
; APPLICANT: Biosynexus Incorporated  
; APPLICANT: Foster, Simon  
; APPLICANT: Mond, James  
; TITLE OF INVENTION: Antigenic Polypeptides  
; FILE REFERENCE: FI00629WO  
; CURRENT APPLICATION NUMBER: US/10/485,517  
; CURRENT FILING DATE: 2004-02-02  
; PRIOR APPLICATION NUMBER: GB 0118925.9  
; PRIOR FILING DATE: 2001-08-02  
; PRIOR APPLICATION NUMBER: GB 0200349.9  
; PRIOR FILING DATE: 2002-01-09  
; NUMBER OF SEQ ID NOS: 424  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 413  
; LENGTH: 270  
; TYPE: PRT  
; ORGANISM: Staphylococcus epidermidis  
US-10-485-517-413

Query Match 46.8%; Score 50.5; DB 6; Length 270;  
Best Local Similarity 55.0%; Pred. No. 1.6;  
Matches 11; Conservative 2; Mismatches 6; Indels 1; Gaps 1;  
QY 2 GGYGHVAVYTVGGQ-GQIQV 20  
DB 222 GPYGHVAVYVERINGDGSILI 241

RESULT 8  
US-10-793-626-1682  
; Sequence 1682, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626  
; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1682  
; LENGTH: 149  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-1682

Query Match 42.1%; Score 45.5; DB 6; Length 149;  
Best Local Similarity 47.6%; Pred. No. 4.6;  
Matches 10; Conservative 3; Mismatches 7; Indels 1; Gaps 1;  
QY 1 DGGYGHVAVYTVGGQ-GQIQV 20  
DB 106 EGAFGHVAFVESVNDGSITV 126

RESULT 9  
US-10-793-626-2870  
; Sequence 2870, Application US/10793626  
; Publication No. US20050255478A1  
; GENERAL INFORMATION:  
; APPLICANT: KIMMERLY, WILLIAM JOHN  
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS  
; FILE REFERENCE: PU3480US  
; CURRENT APPLICATION NUMBER: US/10/793,626

; CURRENT FILING DATE: 2004-03-04  
; PRIOR APPLICATION NUMBER: 60/164,258  
; PRIOR FILING DATE: 1999-11-09  
; NUMBER OF SEQ ID NOS: 4472  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2870  
; LENGTH: 157  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic  
; OTHER INFORMATION: amino acid sequence  
US-10-793-626-2870

Query Match 42.1%; Score 45.5; DB 6; Length 157;  
Best Local Similarity 47.6%; Pred. No. 4.8;  
Matches 10; Conservative 3; Mismatches 7; Indels 1; Gaps 1;  
QY 1 DGGYGHVAVYTVGGQ-GQIQV 20  
DB 106 EGAFGHVAFVESVNDGSITV 126

RESULT 10  
US-10-467-657-8284  
; Sequence 8284, Application US/10467657  
; Publication No. US20050260581A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; APPLICANT: FONTANA Maria Rita  
; APPLICANT: PIZZA Mariagrazia  
; APPLICANT: MASIGNANI Vega  
; APPLICANT: MONACI Elisabetta  
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/467,657  
; CURRENT FILING DATE: 2003-08-11  
; PRIOR APPLICATION NUMBER: GB-0103424.8  
; PRIOR FILING DATE: 2001-02-12  
; NUMBER OF SEQ ID NOS: 9218  
; SOFTWARE: SeqWin99, version 1.04  
; SEQ ID NO 8284  
; LENGTH: 344  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-10-467-657-8284

Query Match 40.7%; Score 44; DB 6; Length 344;  
Best Local Similarity 61.5%; Pred. No. 18;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 4 YGHVAVYTVGGQ 16  
DB 40 YGHAAPFEGGGGG 52

RESULT 11  
US-10-467-657-4224  
; Sequence 4224, Application US/10467657  
; Publication No. US20050260581A1  
; GENERAL INFORMATION:  
; APPLICANT: CHIRON SpA  
; APPLICANT: FONTANA Maria Rita  
; APPLICANT: PIZZA Mariagrazia  
; APPLICANT: MASIGNANI Vega  
; APPLICANT: MONACI Elisabetta  
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/10/467,657  
; CURRENT FILING DATE: 2003-08-11  
; PRIOR APPLICATION NUMBER: GB-0103424.8  
; PRIOR FILING DATE: 2001-02-12  
; NUMBER OF SEQ ID NOS: 9218

```
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 4224
; LENGTH: 1234
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-4224

Query Match      40.3%; Score 43.5; DB 6; Length 1234;
Best Local Similarity 36.0%; Pred. No. 80;
Matches 9; Conservative 5; Mismatches 6; Indels 5; Gaps 1;

QY      1 DGGYGHVAVYVT-----GVQGGQIQV 20
      ||| ||| ||| ||| ||| ||| ||| |||
Db      142 DAGYHNVSHVVAAGFAVRGGIVDL 166

RESULT 12
US-11-055-822-84
; Sequence 84, Application US/11055822
; Publication No. US20050260707A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Kroger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauser, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING
; FILE REFERENCE: BGI-121CPCN
; CURRENT APPLICATION NUMBER: US/11/055,822
; CURRENT FILING DATE: 2005-02-11
; PRIOR APPLICATION NUMBER: 09/606,740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141,031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/142,101
; PRIOR FILING DATE: 1999-07-02
; PRIOR APPLICATION NUMBER: 60/148,613
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: 60/187,970
; PRIOR FILING DATE: 2000-03-09
; PRIOR APPLICATION NUMBER: DE 19930476.9
; PRIOR FILING DATE: 1999-07-01
; PRIOR APPLICATION NUMBER: DE 19931415.2
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931418.7
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931419.5
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1158
; SEQ ID NO 84
; LENGTH: 183
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-11-055-822-84

Query Match      39.8%; Score 43; DB 7; Length 183;
Best Local Similarity 47.1%; Pred. No. 13;
Matches 8; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY      4 YGHVAVYVTGQGQIQV 20
      :|: :|||: |||: |||:
Db      55 HGNNGHVTLTGQIRV 71

RESULT 13
US-11-052-554A-165
; Sequence 165, Application US/11052554A
; Publication No. US20050288866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 165
; LENGTH: 639
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis H37Rv
US-11-052-554A-165

Query Match      39.8%; Score 43; DB 7; Length 639;
Best Local Similarity 56.2%; Pred. No. 48;
Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY      1 DGGYGHVAVYVTGQGG 16
      ||| ||| ||| ||| ||| ||| |||
Db      454 DGGQGGVLYGNGGNGG 469

RESULT 14
US-10-467-657-8694
; Sequence 8694, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 8694
; LENGTH: 756
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-8694

Query Match      39.4%; Score 42.5; DB 6; Length 756;
Best Local Similarity 40.9%; Pred. No. 68;
Matches 9; Conservative 5; Mismatches 5; Indels 3; Gaps 1;

QY      1 DGGYGHV---AYVTGQGGQIQ 19
      |||: |||: |||: |||: |||:
Db      687 DGGFGSLGAAAFVTAQAGRT 708

RESULT 15
US-10-467-657-6872
; Sequence 6872, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON SpA
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
```

```
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqMan9, version 1.04
; SEQ ID NO 6872
; LENGTH: 204
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-6872

Query Match      38.9%; Score 42; DB 6; Length 204;
Best Local Similarity 61.5%; Pred. No. 21;
Matches      8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy      3 GYGHVAYVTGVQG 15
      |:|:|:|:|:|
Db      167 GFGFVYVRGVNG 179

Search completed: February 10, 2006, 23:25:58
Job time : 2.86694 secs
```

**THIS PAGE BLANK (USPTO)**